Multimedia Systems
Lecture – 1

By

Dr. Priyambada Subudhi

Assistant Professor

IIIT Sri City

What is Multimedia?

- The term multimedia is composed of two words
  - Multi : numerous or multiple
  - media : agent for something

- Multimedia could be defined as the usage of multiple media or agents for communication of information.

- These media or agents could be in the form of text, images, audio, video, graphics, animation etc.

- A good general working definition is

Multimedia is the field concerned with the computer controlled integration of text, graphics, drawings, still and moving images (Video), animation, audio, and any other media where every type of information can be represented, stored, transmitted and processed digitally.

- Example: A music video and sound should be used together as one without another would lose its significance.

- The term media can be categorized based on few criteria
  - Perception media
  - Representation media
  - Presentation media
  - Storage media
  - Transmission media

Perception media

- “How do human perceive information”
- We perceive information from what we see and what we hear.
- Visual media
  - Text, graphics, images and video
- Auditory media
  - Music, sound, voice

Representation media

- "How information is encoded in the computer"
- Referring to how the information is represented internally in computer.
- Several options:
  - Text is encoded in ASCII
  - An audio data stream in PCM
  - Image in JPEG format
  - Video in MPEG format

Presentation media

- "Which medium is used to output information from the computer or input in the computer"
- Refers to physical means used by systems to reproduce information for humans, e.g: audio and visual devices.
- Input
  - Keyboards, cameras, microphone, Head mounted device
- Output
  - Paper, monitors, loud speakers

Storage media

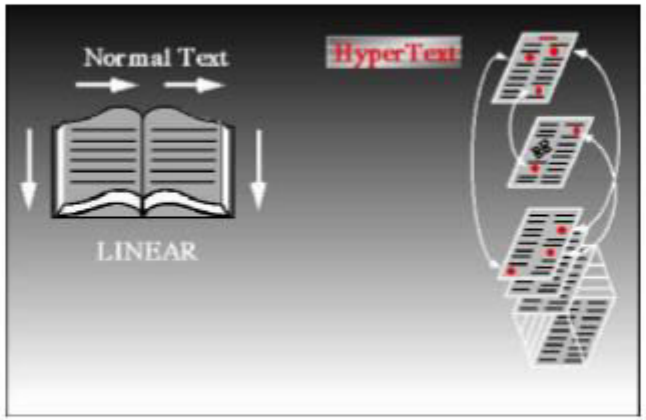
- "Where information is stored"
- Refers to various physical means for storing computer data such as magnetic tapes, magnetic disks or digital optical disks (CD-ROM, CD, DVD)

Transmission media

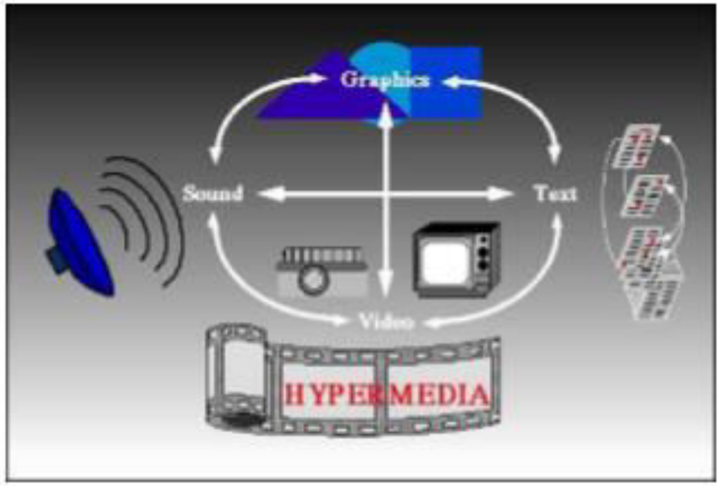
- "Which medium is used to transmit data"
- Refers to physical means – cable of various type (coaxial cable, twisted pair, fibre optics), radio tower, satellite – that allow the transmission of telecommunication signals.

Characteristics of multimedia

- It can increase the impact of the message or impact on the user.
- Interactivity:
  - When the end-user is able to control the elements of media that are required, and subsequently obtains the required information in a non-linear way.
- Hypertext and Hypermedia support:
  - Hypertext is a text which contains links to other texts.
  - Traversal through pages of hypertext is therefore usually non-linear (as indicated below)



- HyperMedia is not constrained to be text-based. It can include other media, e.g., graphics, images, and especially continuous media – sound and video.
- The World Wide Web is the largest and most commonly used hypermedia application.



- It can involve more than one input device.
- It can repeated (over and over).
- It is generally dynamic, not static.

Categories of Multimedia

- Multimedia may be divided into following three categories based on their functions and how they are organized.
- Linear and non-linear: Linear active content progresses without any navigation control for the viewer such as a cinema presentation. Non-linear content offers user interactivity to control progress as used with a computer game or used in self-paced computer-based training.
- Interactive and non-interactive: Interactive multimedia is the means to interface with different media through input (e.g., a computer keyboard, mouse, touch screen, on screen buttons, and text entry, etc.) and output devices allowing a user to make decisions as to what takes place next with multimedia
- Real-time and recorded: Multimedia presentations can be live (real-time) or recorded. A recorded presentation may allow interactivity via a navigation system. A live multimedia presentation may allow interactivity via interaction with the presenter or performer.

Multimedia Applications

- **Definition**: A Multimedia Application is an application which uses a collection of multiple media sources e.g. text, graphics, images, sound/audio, animation and/or video.
- **Examples:**
  - **Video Teleconferencing:** Transmission of synchronized video and audio in real-time through computer networks in between two or more multipoints (or participants) separated by locations.
  - **Multimedia Store and Forward Mail**: Allow users to generate, modify and receive documents that contain multimedia. Eg. Gmail, Hotmail, Yahoo etc.
  - **Advertising and Purchasing:** Most of the web sites visited have many advertisements with multimedia features with the objective of marketing merchandise or offering services online.

- For entertainment:
  - Multimedia is heavily used in the entertainment industry, especially to develop special effects in movies and animations. Multimedia games are a popular pastime and are software programs available either as CD-ROMs or online
- For Education:
  - Multimedia is used to produce computer-based training courses (popularly called CBTs) and reference books like encyclopedia and almanacs.
- For Healthcare:
  - Multimedia best use in healthcare is for real time monitoring of conditions of patients in critical illness or accident.
  - Multimedia makes it possible to consult a surgeon or an expert who can watch an ongoing surgery line on his PC monitor and give online advice at any crucial juncture.
- Other multimedia applications available to us at home are
  - Basic Television Services, Digital Audio, Video on demand, Home shopping, Digital multimedia libraries, E-Newspapers, e-magazines

History of Multimedia

- Multimedia to communicate ideas might begin with newspapers, which were perhaps the first mass communication medium, using text, graphics, and images.
- Thomas Alva Edison 'commissioned the invention of a motion picture camera in 1887.
- In 1895, Guglielmo Marconi sent his first wireless radio transmission at POlltecchio, Italy. Initially invented for telegraph, radio is now a major medium for audio broadcasting.
- Television was the new medium for the twentieth century. It established video as a commonly available medium and has since changed the world of mass communication.

The connection between computers and digital media represented using the discrete binary format, emerged only over a short period:

1945 : As part of MIT's postwar deliberations on what to do with all those scientists employed on the war effort, Vannevar Bush (1890-1974) wrote a landmark article describing what amounts to a hypermedia system, called "Memex."

1960s Ted Nelson started the Xanadu project and coined the term "hypertext."

1967 Nicholas Negroponte formed the Architecture Machine Group at MIT.

1969 Nelson and van Dam at Brown University created an early hypertext editor called FRESS.

1976 The MIT Architecture Machine Group proposed a project entitled "Multiple Media." This resulted in the Aspen Movie Map, the first videodisk, in 1978.

1982 The Compact Disc (CD) was made commercially available by Philips and Sony, which was soon becoming the standard and popular medium for digital audio data.

1985 Negroponte and Wiesner co-founded the MIT Media Lab, a leading research institution investigating digital video and multimedia.

1990 Kristina Hooper Woolsey headed the Apple Multimedia Lab, with a staff of 100.

1991 MPEG1 was approved as an international standard for digital video. Its further development led to newer standards, MPEG-2, MPEG- 4, and further MPEGs, in the 1990s.

1991 The introduction of PDAs in 1991 began a new period in the use of computers in general and multimedia in particular.

1992 JPEG was accepted as the international standard for digital image compression, which remains widely used today.

1992 The first audio multicast on the multicast backbone (MBone) was made.

1995 The JAVA language was created for platform-independent application development, which was widely used for developing multimedia applications.

1996 DVD video was introduced; high-quality, full-length movies were distributed on a single disk.

1998 Handheld MP3 audio players were introduced to the consumer market, initially with 32 MB of flash memory.

2000 World Wide Web (WWW) size was estimated at over 1 billion pages.

2001 The first peer-to-peer file sharing (mostly MP3 music) system.

2003 Skype was released for free peer-to-peer voice over the Internet.

2004 Web 2.0 was recognized as a new way to utilize software developers and end-users use the Web.

2005 YouTube was created, providing an easy portal for video sharing, which was purchased by Google in late 2006.

2006 Twitter was created, and rapidly gained world wide popularity, with 500 million registered users in 2012.

2007 Apple launched the first generation of iPhone, running the iOS mobile operating system.

2008 The first Android-powered phone was sold and Google Play the Android's primary app store, was soon launched.

2009 The first LTE (Long Term Evolution) network was setup making an important step toward 4G wireless networking.

2010 Netflix, which used to be a DVD rental service provider, migrated its infrastructure to the Amazon AWS cloud computing platform, and became a major online streaming video provider.

Multimedia Systems
Lecture – 2

By

Dr. Priyambada Subudhi
Assistant Professor
IIIT Sri City

Elements of Multimedia Data

- The common elements of multimedia includes
  - Text: All multimedia productions contain some amount of text. The text can have various types of fonts and sizes to suit the profession presentation of the multimedia software.
  - Graphics:
    - Graphics make the multimedia application attractive.
    - In many cases people do not like reading large amount of textual matter on the screen. Therefore, graphics are used more often than text to explain a concept, present background information etc.
    - There are two types of Graphics:
      - Bitmap Images : Bitmap images are real images that can be captured from devices such as digital cameras or scanners. Generally bitmap images are not editable. Bitmap images require a large amount of memory.
      - Vector Graphics : Vector graphics are drawn on the computer and only require a small amount of memory. These graphics are editable.

- Audio:
    - A multimedia application may require the use of speech, music and sound effects. These are called audio or sound element of multimedia.
    - Speech is also a perfect way for teaching. Audio are of analog and digital types. Analog audio or sound refers to the original sound signal.
    - Computer stores the sound in digital form. Therefore, the sound used in multimedia application is digital audio.
- Video:
    - The term video refers to the moving picture, accompanied by sound such as a picture in television.
    - Video element of multimedia application gives a lot of information in small duration of time.
    - Digital video is useful in multimedia application for showing real life objects.
    - Video have highest performance demand on the computer memory and on the bandwidth if placed on the internet.
    - Digital video files can be stored like any other files in the computer.

Animation:

- Animation is a process of making a static graphical elements look like it is moving.
- An animation is just a continuous series of still graphical elements that are displayed in a sequence.
- The animation can be used effectively for attracting attention.
- Animation also makes a presentation light and attractive.

Multimedia Data Representation: Text

- Source: Keyboard, speech input, optical character recognition, data stored on disk
- Stored and input character by character:
  - Storage of text is 1 byte per char / more bytes for Unicode.
  - For other forms of data (e.g. Spreadsheet les). May store format as text (with formatting) others may use binary encoding.
- Format: Raw text or formatted text e.g HTML, Rich Text Format (RTF), Word or a program language source (Java, Python, MATLAB etc.)
- Not temporal. But may have natural implied sequence e.g. HTML format sequence, Sequence of C program statements.
- Size Not significant w.r.t. other Multimedia data.

Images

- Input: digitally scanned photographs/pictures or directly from a digital camera.
- May also be generated by programs "similar" to graphics or animation programs.
- Still pictures which (uncompressed) are represented as a bitmap (a grid of pixels organized as a 2D array).
- The two dimensions specify the width and height of the images. Each pixel has also a bit depth which represents the number of bits assigned to each pixel.
- Stored at 1 bit per pixel (Black and White), 8 Bits per pixel (Grey Scale, Colour Map) or 24 Bits per pixel (True Colour)
- Size: a 512x512 Gray scale image takes up 1/4 MB, a 512x512 24 bit image takes 3/4 MB with no compression.
- This overhead soon increases with image size.

Graphics

- Format: constructed by the composition of primitive objects such as lines, polygons, circles, curves and arcs.
- Input: Graphics are usually generated by a graphics editor program (e.g. Illustrator) or automatically by a program (e.g. Postscript).
- Graphics are usually editable or revisable (unlike Images).
- Graphics input devices: keyboard (for text and cursor control), mouse, trackball or graphics tablet.
- Graphics standards : OpenGL, PHIGS, GKS
- Graphics files usually store the primitive assembly
- Do not take up a very high storage overhead.

Audio

- Digital audio is characterized by a sampling rate in hertz, which gives the number of samples per second.
- A sample can be defined as an individual unit of audio information.
- Each sample also has a size, the sample size, which typically is anywhere from 8-bits to 16-bits depending on the application.
- CD Quality Audio requires 16-bit sampling at 44.1 KHz Even higher audiophile rates (e.g. 24-bit, 96 KHz)
- Audio signal is also described by dimensionality i.e. the dimensions of an audio signal signify the number of channels that are contained in the signal. These may be mono (one channel), stereo (two channels).
- 1 Minute of Mono CD quality (uncompressed) audio requires 5 MB.
- 1 Minute of Stereo CD quality (uncompressed) audio requires 10 MB.
- Usually compressed (E.g. MP3, AAC, Flac, Ogg Vorbis).

Video

- Input: Analog Video is usually captured by a video camera and then digitized.
- Video is represented as a sequence of images. Each image in the sequence typically has the same properties of width, height, and pixel depth.
- There is one more temporal parameter known as frames per second or fps.
- Typically, videos have 25, 30 or 50 frames per second.
- E.g. A 512×512 size monochrome video images take 25×0.25 = 6.25MB for a second to store uncompressed.
- Typical PAL digital video (720×576 pixels per colour frame) 1.24 × 25 = 31MB for a second to store uncompressed.
- High Definition video on Blu-ray (up to 1920×1080 = 2 Megapixels per frame) 6.2×25 = 155MB for a second to store uncompressed. (There are higher possible frame rates!)
- Digital video clearly needs to be compressed for most times.

Multimedia Systems
Lecture – 3

By

Dr. Priyambada Subudhi
Assistant Professor
IIIT Sri City

Multimedia Software Tools

- For a concrete appreciation of the current state of multimedia software tools available for carrying out tasks in multimedia.
- The following categories of software tools we examine here
  - **Music Sequencing and Notation:** e.g. Cakewalk Pro Audio, Finale, Sibelius
  - **Digital Audio:** Digital Audio tools deal with accessing and editing the actual sampled sounds that make up audio.

    e.g. Adobe Audition, Sound Forge, Pro Tools
  - **Graphics and Image Editing:** e.g. Adobe Illustrator, Adobe Photoshop, Adobe Fireworks, Adobe Freehand

- **Video Editing tools:** e.g. Adobe Premiere, CyberLink PowerDirector, Adobe After Effects, Final Cut Pro
- **Animation:**
  - **Multimedia APIs:** Java3D, DirectX, OpenGL
  - **Animation Software**: Autodesk 3ds Max, Autodesk Softimage, Autodesk Maya
  - **GIF Animation Packages**
- **Multimedia Authoring:** Tools that provide the capability for creating a complete multimedia presentation, including interactive user control, are called *authoring* programs.

  E.g. Adobe Flash, Adobe Director, Dreamweaver

Multimedia Systems

- A Multimedia System is a system capable of processing multimedia data and applications.

- A Multimedia System is characterized by the processing, storage, generation, manipulation and rendition of Multimedia information.

Characteristics of Multimedia Systems

- A Multimedia system has four basic characteristics:
  - Multimedia systems must be computer controlled.
  - Multimedia systems are integrated.
  - The information they handle must be represented digitally.
  - The interface to the final presentation of media is usually interactive.

Challenges for Multimedia Systems

- Distributed Networks
- Temporal relationship between data
  - Render different data at same time- continuously.
  - Sequencing within the media:

    playing frames in correct order/time frame in video
  - Synchronisation- inter-media scheduling

e.g. Video and Audio - Lip synchronization is clearly important for humans to watch playback of video and audio and even animation and audio.

Key Issues for Multimedia Systems

The key issues multimedia systems need to deal with here are:

- How to represent and store temporal information.
- How to strictly maintain the temporal relationships on play back/retrieval
- What process are involved in the above.
- Data has to represented digitally- Analog-Digital Conversion, Sampling etc.
- Large Data Requirements- bandwidth, storage

  Data compression is usually mandatory

Desirable Features for a Multimedia System

Given the above challenges the following feature a desirable (if not a prerequisite) for a Multimedia System:

- Very High Processing Power- needed to deal with large data processing and real time delivery of media. Special hardware commonplace.
- Multimedia Capable File System- needed to deliver real-time media -e.g. Video/Audio Streaming.
- Special Hardware/Software needed - e.g. RAID technology.
- Data Representations - File Formats that support multimedia should be easy to handle yet allow for compression/decompression in real-time.

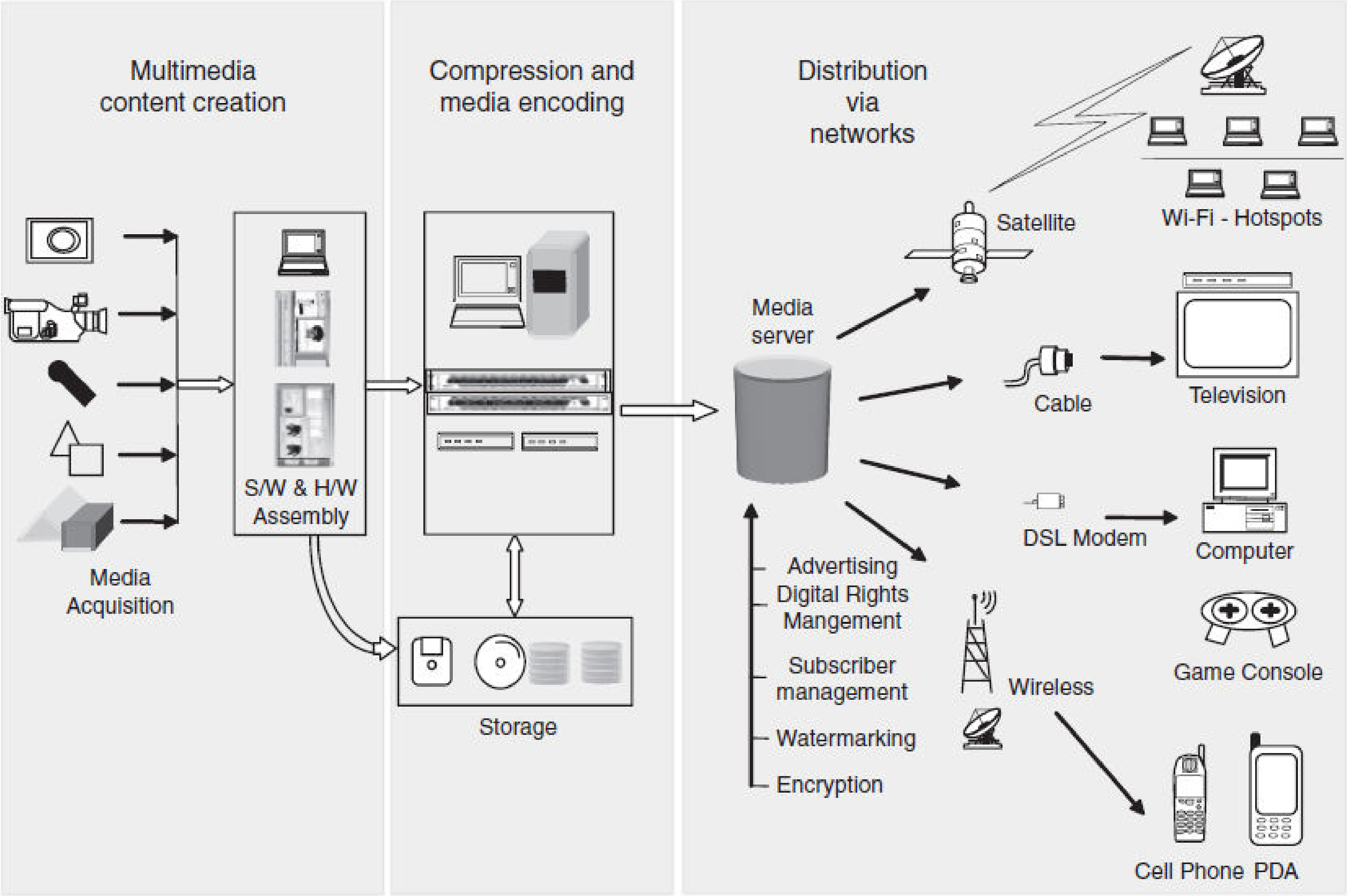
- Efficient and High I/O – input and output to the file subsystem needs to be efficient and fast. Needs to allow for real-time recording as well as playback of data. e.g. Direct to Disk recording systems.
- Special Operating System - to allow access to file system and process data efficiently and quickly. Needs to support direct transfers to disk, real-time scheduling, fast interrupt processing, I/O streaming etc.
- Storage and Memory - large storage units (of the order of hundreds of Tb if not more) and large memory (several Gb or more). Large Caches also required and high speed buses for efficient management.
- Network Support - Client-server systems common as distributed systems common.
- Software Tools - user friendly tools needed to handle media, design and develop applications, deliver media.

Components of Multimedia Systems

Multimedia systems can be logically grouped into three parts whose primary functionalities are

- Content production
- Compression and storage
- Distribution to various end users and platforms

Components of a multimedia system today



- Content Production:
  - It includes a variety of different instruments, which capture different media types in a digital format. These include digital cameras, camcorders or video cameras, sound recording devices, scanners to scan images, and 3D graphical objects.
  - Once the individual media elements are in their digital representations, they may be further combined to create coherent, interactive presentations using software (S/W) applications and hardware (H/W) elements.
  - This content can be stored to disk, or in the case of real-time applications, the content can be sent directly to the end user via digital networks.

- Compression and Storage:
  - It deals with the compression of multimedia content
  - This entails the use of various compression technologies to compress video, audio, graphics, and so on.
- Distribution
  - It deals with media distribution across a variety of low-bandwidth and high-bandwidth networks.
  - This ranges from cellular, to wireless networks, to cable, to digital subscriber line (DSL), to satellite networks.
  - Distribution normally follows standards protocols, which are responsible for collating and reliably sending information to end receivers.
  - The commonly used end receivers are computers, televisions, set-top boxes, cell phones, or even more application- or entertainment-specific items, such as video game consoles.

Multimedia Systems Lecture – 4

By

Dr. Priyambada Subudhi
Assistant Professor
IIIT Sri City

Text: Visual Representation

- Text has become a part of our life. It consists of characters, punctuation symbols, etc. to convey a message.
- Text is one of the most imperative components of multimedia and an essential source of presenting information to a wide range of people.
- Proper use of text, keeping in mind elements such as font style, size and various design tools help the content creator to communicate the idea and message to the user.

Fonts and Faces

- Factors affecting legibility of text are as follows:
  - Size and style
  - Background and foreground colours
  - Leading
- A glyph is a graphic representation of a character's shape where a character may be represented by many glyphs.
- A typeface is the design of lettering that include variations in size, weight, slope, width etc.
- A font is a collection of character or glyphs of a single size and style belonging to a particular typeface family.
- Fonts are classified on the basis of spacing between characters, words, presence or absence of serifs, their shape, stretch and weight such as bold or italics.

- The spacing between character pairs is called kerning and the space between lines is called leading.

Figure 2.1: Different Types of Fonts

Arial

Arial Black

Comic Sans MS

Courier New

Georgia

Impact

Times New Roman

- Font size is measured in points and it does not describe the height or width of its character. This happens because the height of two different fonts (in both upper and lower case) may differ.
- One point is approximately 1/72 of an inch i.e., 0.0138.

Figure 2.2: Examples of Different Fonts and Font Sizes

There are three main stages of a multimedia project.

Lucida Handwriting 12 point

There are three main stages of a multimedia project.

Kristen ITC 10 point

THERE ARE THREE MAIN STAGES OF A MULTIMEDIA PROJECT.

Big Truck 18 point

There are three main stages of a multimedia project.

Microsoft Sans 8 point

Example: Times New Roman, Bodoni, Bookman are some fonts which come under serif category.

Arial, Avant Garde, Verdana are some examples of sans serif font.

Figure 2.3: Examples of Serif, Sans Serif and Decorative Fonts

Bodoni
Interactive multimedia is called hypermedia.
(This is a serif font. In this font, a line or curve extension from the end of a letter. Serif fonts are best used for body text.)

Avant Garde
Interactive multimedia is called hypermedia.
(This is a sans serif font. There are no extensions in this font. Sans Serif fonts are best used for titles.)

Matura MT Script
Interactive multimedia is called hypermedia.
(This is a decorative font. These fonts are stylish and formal and are best used for emphasis.)

Selecting Text Fonts

There are a few things that a user must keep in mind before selecting fonts for a multimedia presentation.

- Choose a font that is legible and easy to read.
- The different effects and colours of a font can be chosen to make the text look distinctive.
- Try to use few different colours within the same presentation.
- Try to use few typefaces within the same presentation. Play with the style and size to match up to the purpose and importance of the text.
- Drop caps and initial caps can be used to accent the words.
- To attract instant attention to the text, the words can be wrapped onto a sphere or bent like a wave.
- In case of text links (anchors) on web pages the messages can be highlighted.
- Meaningful words and phrases can be used for links and menu items.
- Overcrowding of text on a single page should be avoided.
- Do not use decorative passages for longer paragraphs.

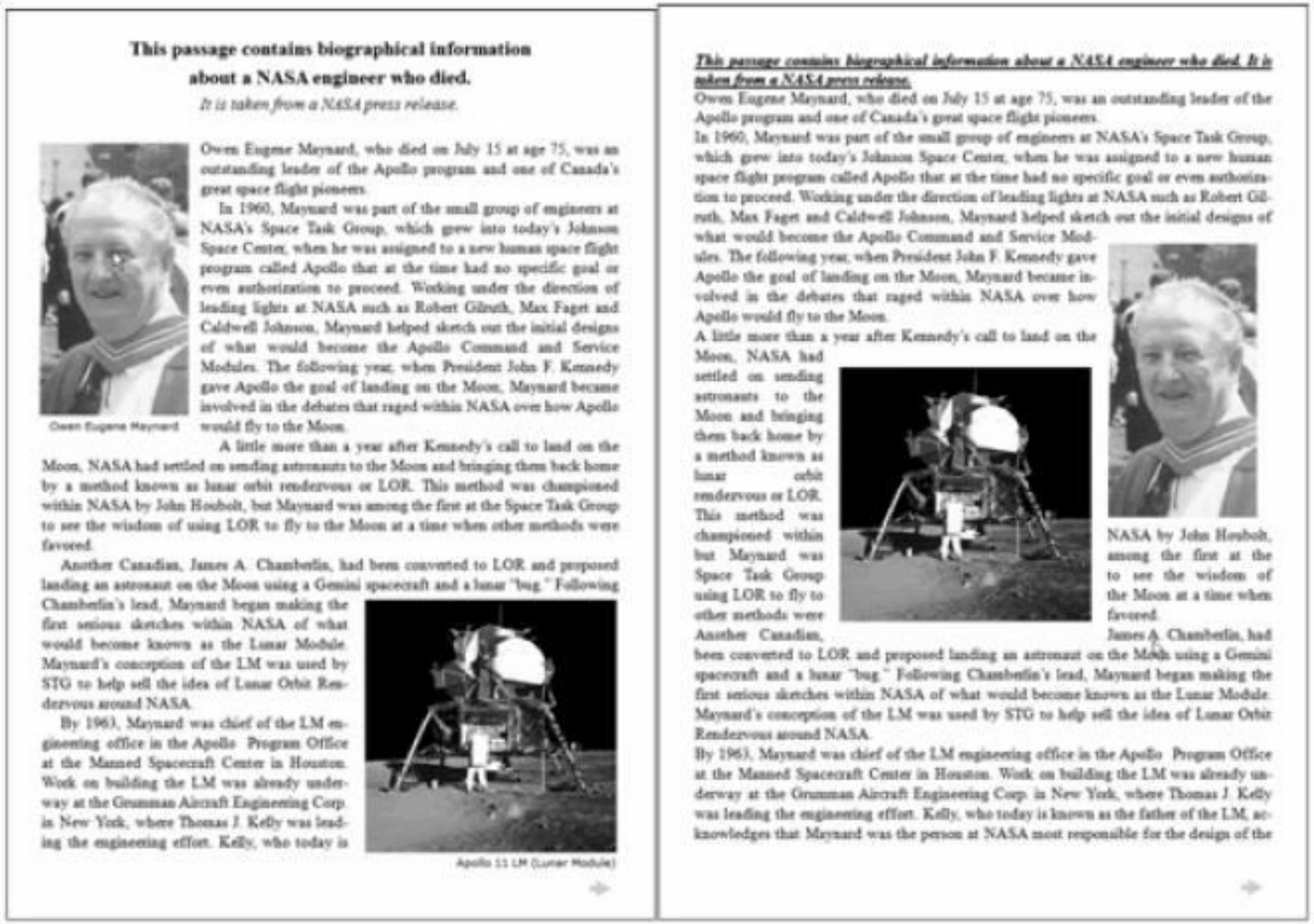
Using Text Elements in a Multimedia Presentation

The text elements used in multimedia are given below:

- **Menus for Navigation**
  - A user navigates through content using a menu.
  - A simple menu consists of a text list of topics.
- **Interactive Buttons**
  - A button is a clickable object that executes a command when activated.
  - Users can create their own buttons from bitmaps and graphics.
  - The design and labelling of the buttons should be treated as an industrial art project.
- **Fields for Reading**
  - Reading a hard copy is easier and faster than reading from the computer screen.
  - A document can be printed in one of two orientations - portrait or landscape.

- **HTML Documents**
  - HTML stands for Hypertext Markup Language which is the standard document format used for Web pages.
  - HTML documents are marked using tags.
  - Some of the commonly used tags are:
    - The <B> tag for making text bold faced.
    - The <OL> tag for creating an ordered list.
    - The <IMG> tag for inserting images.
- **Text Layout**
  - While creating a multimedia presentation, the presenter should plan the text layout to let a reader read it with ease.
  - The length of the text should neither too long nor too short.
  - For a printed document, a line containing 13 to 17 words is sufficient.
  - The next point of concern is the consistency of pages. The designer has to make sure that the pages should be of same size.

Figure 2.4: Example of Good and Poor Page Layout



This passage contains biographical information about a NASA engineer who died.

It is taken from a NASA press release.

Owen Eugene Maynard

Owen Eugene Maynard, who died on July 15 at age 75, was an outstanding leader of the Apollo program and one of Canada's great space flight pioneers.

In 1960, Maynard was part of the small group of engineers at NASA's Space Task Group, which grew into today's Johnson Space Center, when he was assigned to a new human space flight program called Apollo that at the time had no specific goal or even authorization to proceed. Working under the direction of leading lights at NASA such as Robert Gilruth, Max Faget and Caldwell Johnson, Maynard helped sketch out the initial designs of what would become the Apollo Command and Service Modules. The following year, when President John F. Kennedy gave Apollo the goal of landing on the Moon, Maynard became involved in the debates that raged within NASA over how Apollo would fly to the Moon.

A little more than a year after Kennedy's call to land on the Moon, NASA had settled on sending astronauts to the Moon and bringing them back home by a method known as lunar orbit rendezvous or LOR. This method was championed within NASA by John Houbolt, but Maynard was among the first at the Space Task Group to see the wisdom of using LOR to fly to the Moon at a time when other methods were favored.

Another Canadian, James A. Chamberlin, had been converted to LOR and proposed landing an astronaut on the Moon using a Gemini spacecraft and a lunar "bug." Following Chamberlin's lead, Maynard began making the first serious sketches within NASA of what would become known as the Lunar Module. Maynard's conception of the LM was used by STG to help sell the idea of Lunar Orbit Rendezvous around NASA.

By 1963, Maynard was chief of the LM engineering office in the Apollo Program Office at the Manned Spacecraft Center in Houston. Work on building the LM was already underway at the Grumman Aircraft Engineering Corp. in New York, where Thomas J. Kelly was leading the engineering effort. Kelly, who today is

Apollo 11 LM (Lunar Module)

This passage contains biographical information about a NASA engineer who died. It is taken from a NASA press release.

Owen Eugene Maynard, who died on July 15 at age 75, was an outstanding leader of the Apollo program and one of Canada's great space flight pioneers.

In 1960, Maynard was part of the small group of engineers at NASA's Space Task Group, which grew into today's Johnson Space Center, when he was assigned to a new human space flight program called Apollo that at the time had no specific goal or even authorization to proceed. Working under the direction of leading lights at NASA such as Robert Gilruth, Max Faget and Caldwell Johnson, Maynard helped sketch out the initial designs of what would become the Apollo Command and Service Modules. The following year, when President John F. Kennedy gave Apollo the goal of landing on the Moon, Maynard became involved in the debates that raged within NASA over how Apollo would fly to the Moon.

A little more than a year after Kennedy's call to land on the Moon, NASA had settled on sending astronauts to the Moon and bringing them back home by a method known as lunar orbit rendezvous or LOR. This method was championed within but Maynard was Space Task Group using LOR to fly to other methods were Another Canadian, NASA by John Houbolt, among the first at the to see the wisdom of the Moon at a time when favored. James A. Chamberlin, had been converted to LOR and proposed landing an astronaut on the Moon using a Gemini spacecraft and a lunar "bug." Following Chamberlin's lead, Maynard began making the first serious sketches within NASA of what would become known as the Lunar Module. Maynard's conception of the LM was used by STG to help sell the idea of Lunar Orbit Rendezvous around NASA.

By 1963, Maynard was chief of the LM engineering office in the Apollo Program Office at the Manned Spacecraft Center in Houston. Work on building the LM was already underway at the Grumman Aircraft Engineering Corp. in New York, where Thomas J. Kelly was leading the engineering effort. Kelly, who today is known as the father of the LM, acknowledges that Maynard was the person at NASA most responsible for the design of the

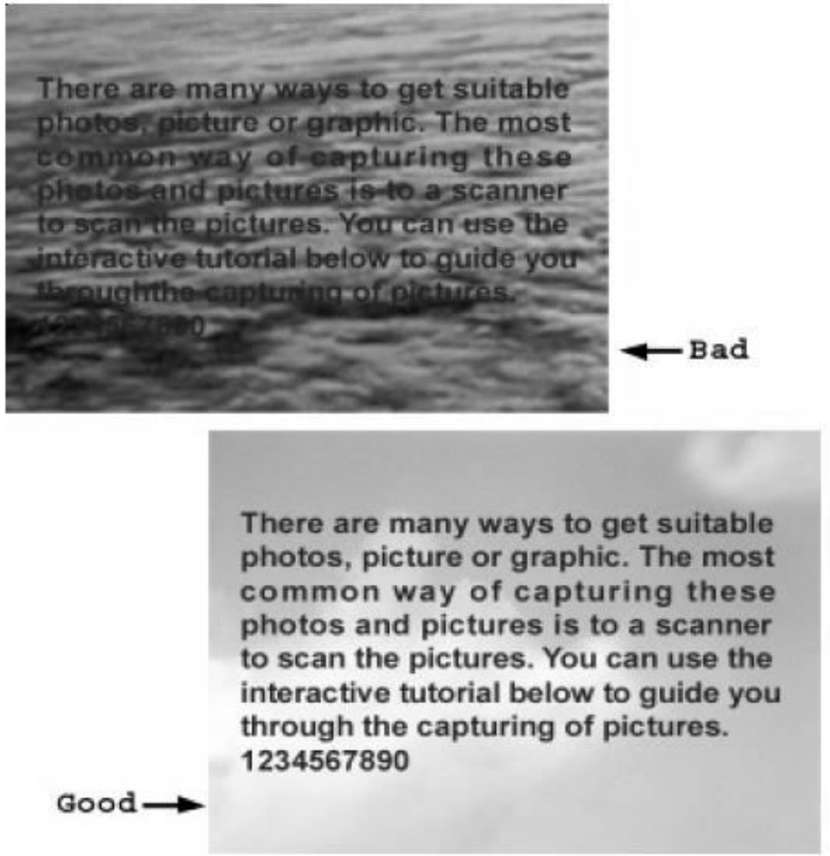
- The distance between two lines should be adjusted to a suitable value to increase readability.
- Ensure that the leading is not too small as then the text will be hard to read.

Figure 2.5: Effects of Different Leadings

Tight	If your software allows you to change the leading (the distance between two text lines), you should adjust it to a suitable value.
Normal	If your software allows you to change the leading (the distance between two text lines), you should adjust it to a suitable value.
Loose	If your software allows you to change the leading (the distance between two text lines), you should adjust it to a suitable value.

- The other one of the most common errors people make is while choosing the background colour. Using a background colour too close to the text or a background image highly in contrast to the text which makes the text difficult to read.

Figure 2.6: Use of Colour and Background



Use of Text in Webs

- Using text in websites attract a visitor's attention as well as help him in understanding the webpage better.
- It is far better than the use of meaningless graphics and images which do not contribute in understanding of the page.
- **Website Loading Speed:** A website which contains a lot of text loads faster than the websites that contains a lot of images and graphics, audio and video clips on the page.

Text in Films

- Most films start with titles and end with credits.
- Typography look different in different formats such as a in film subtitles, on websites, poster etc.
- While designing subtitles, a film maker will need to keep in mind that moving images interact with the top layer subtitles.
  - E.g. If subtitles are white and rest on top of a similar white tone in the image, the text will be difficult and impossible to read. To ensure this does not happen, a black outline around text should be used. Now the text will be viewable even against common black and white backgrounds.

Text in Advertisements

- Since the text ads are more of keyword oriented, they draw more attention than banner advertising.
- The text ads are inexpensive, thus making it affordable and effective for your business.
- There are a few websites which offers a flat free rental services to place your text based advertisements.
- The foremost benefit of having text based advertisements is that it helps in improving your search engine ranking.

Multimedia Systems
Lecture – 5

By

Dr. Priyambada Subudhi

Assistant Professor

IIIT Sri City

Digital Representation of Text

- Any type of data needs to be converted to its digital form (in the form 1s and 0s) to be represented in the Computer.

- Text (letters, symbols, and numerals) also needs to be encoded in digital form is what allows computers to manipulate and communicate text.

- Two character encoding standards define how characters are decoded from ones and zeros into the text you see on the screen right now, and into the different languages viewed every day on the World Wide Web.

- These two encoding standards are ASCII and Unicode.

ASCII

- The American Standard Code for Information Interchange (ASCII) was developed to create an international standard for encoding the Latin alphabet.
- In 1963, ASCII was adopted so information could be interpreted between computers; representing lower and upper letters, numbers, symbols, and some commands.
- Because ASCII is encoded using ones and zeros, the base 2 number system, it uses seven bits.
- Seven bits allow 2 to the power of 7 = 128 possible combinations of digits to encode a character.

ASCII control characters

00	NULL	(Null character)
01	SOH	(Start of Header)
02	STX	(Start of Text)
03	ETX	(End of Text)
04	EOT	(End of Trans.)
05	ENQ	(Enquiry)
06	ACK	(Acknowledgement)
07	BEL	(Bell)
08	BS	(Backspace)
09	HT	(Horizontal Tab)
10	LF	(Line feed)
11	VT	(Vertical Tab)
12	FF	(Form feed)
13	CR	(Carriage return)
14	SO	(Shift Out)
15	SI	(Shift In)
16	DLE	(Data link escape)
17	DC1	(Device control 1)
18	DC2	(Device control 2)
19	DC3	(Device control 3)
20	DC4	(Device control 4)
21	NAK	(Negative acknowl.)
22	SYN	(Synchronous idle)
23	ETB	(End of trans. block)
24	CAN	(Cancel)
25	EM	(End of medium)
26	SUB	(Substitute)
27	ESC	(Escape)
28	FS	(File separator)
29	GS	(Group separator)
30	RS	(Record separator)
31	US	(Unit separator)
127	DEL	(Delete)

ASCII printable characters

32	space	64	@	96	`
33	!	65	A	97	a
34	"	66	B	98	b
35	#	67	C	99	c
36	$	68	D	100	d
37	%	69	E	101	e
38	&	70	F	102	f
39	'	71	G	103	g
40	(	72	H	104	h
41	)	73	I	105	i
42	*	74	J	106	j
43	+	75	K	107	k
44	,	76	L	108	l
45	-	77	M	109	m
46	.	78	N	110	n
47	/	79	O	111	o
48	0	80	P	112	p
49	1	81	Q	113	q
50	2	82	R	114	r
51	3	83	S	115	s
52	4	84	T	116	t
53	5	85	U	117	u
54	6	86	V	118	v
55	7	87	W	119	w
56	8	88	X	120	x
57	9	89	Y	121	y
58	:	90	Z	122	z
59	;	91	[	123	{
60	<	92	\	124		
61	=	93	]	125	}
62	>	94	^	126	~
63	?	95	_		

Extended ASCII characters

128	Ç	160	á	192	└	224	Ó
129	ü	161	í	193	┴	225	ß
130	é	162	ó	194	┬	226	Ô
131	â	163	ú	195	├	227	Ò
132	ä	164	ñ	196	─	228	õ
133	à	165	Ñ	197	┼	229	Õ
134	å	166	ª	198	ã	230	µ
135	ç	167	º	199	Ã	231	þ
136	ê	168	¿	200	╚	232	Þ
137	ë	169	®	201	╔	233	Ú
138	è	170	¬	202	╩	234	Û
139	ï	171	½	203	╦	235	Ù
140	î	172	¼	204	╠	236	ý
141	ì	173	¡	205	═	237	Ý
142	Ä	174	«	206	╬	238	¯
143	Å	175	»	207	¤	239	´
144	É	176	░	208	ð	240	≡
145	æ	177	▒	209	Ð	241	±
146	Æ	178	▓	210	Ê	242	‗
147	ô	179	│	211	Ë	243	¾
148	ö	180	┤	212	È	244	¶
149	ò	181	Á	213	ı	245	§
150	û	182	Â	214	Í	246	÷
151	ù	183	À	215	Î	247	¸
152	ÿ	184	©	216	Ï	248	°
153	Ö	185	╣	217	┘	249	¨
154	Ü	186	║	218	┌	250	·
155	ø	187	╗	219	█	251	¹
156	£	188	╝	220	▄	252	³
157	Ø	189	¢	221	¦	253	²
158	×	190	¥	222	Ì	254	■
159	ƒ	191	┐	223	▀	255	nbsp

How encoding ASCII works

- You already know how to convert between decimal and binary numbers
- You now need to turn letters into binary numbers
- Every character has a corresponding decimal number (for example, A → 65)
- ASCII uses 7 bits
- We use the first 7 columns of the conversion table to create 128 different numbers (from 0 to 127)

- Example: 1000001 gives us the number 65 (64+1), which corresponds to the letter A

64	32	16	8	4	2	1
1	0	0	0	0	0	1

- Here’s how ‘HELLO’ is encoded in ASCII in binary:

Latin character	ASCII
H	1001000
E	1000101
L	1001100
L	1001100
O	1001111

Let's apply this theory in practice:

- Open Notepad, or whichever plain text editor you prefer
- Type a message and save it, e.g. 'data is beautiful'
- Look at the size of the file — mine is 18 bytes
- Now, add another word, e.g. 'data is SO beautiful'
- If you look at the file size again, you'll see that it has changed — my file is now 3 bytes larger (SO[SPACE]: the 'S', the 'O', and the space)

Extended ASCII

- Because ASCII encodes characters in 7 bits, moving to 8-bit computing technology meant there was one extra bit to be used.
- With this extra digit, **Extended ASCII** encoded up to 256 characters.
- However, the problem that developed was that countries that used different languages did different things with this extra capacity for encoding.
- Many countries added their own additional characters, and different numbers represented different characters in different languages.
- The problem of incompatible encoding systems became more urgent with the invention of the World Wide Web, as people shared digital documents all over the world, using multiple languages.
- To address the issue, the **Unicode Consortium** established a universal encoding system called **Unicode**.

Unicode

- Unicode (Unique, Universal, and Uniform character enCoding) is the new standard for representing characters of all the languages of the World.
- The latest version of Unicode contains a repertoire of more than 120,000 characters covering 129 modern and historic scripts, as well as multiple symbol sets.
- ASCII character encoding is a subset of Unicode.
- Unicode can be implemented by different character encodings. The Unicode standard defines UTF-8, UTF-16 and UTF-32.
- So, these use between 8 and 32 bits per character and has the advantage that it represents many more unique characters than ASCII because of the larger number of bits available to store a character code.
- It uses the same codes as ASCII up to 127.

- UTF-8 the dominant encoding on the World Wide Web (*used in over 92% of websites*), uses one byte for the first 128 code points, and up to 4 bytes for other characters. The first 128 Unicode code points are the ASCII characters, which means that any ASCII text is also a UTF-8 text.
- UTF-16 uses 16bits to represent each character. This means that it is capable of representing 65,536 different characters.
- UTF-32 used 32bits to represent each character, meaning it can represent a character set 4,294, 967,298 possible characters, enough for all known languages.

character	encoding	bits
A	UTF-8	`01000001`
A	UTF-16	`00000000 01000001`
A	UTF-32	`00000000 00000000 00000000 01000001`
あ	UTF-8	`11100011 10000001 10000010`
あ	UTF-16	`00110000 01000010`
あ	UTF-32	`00000000 00000000 00110000 01000010`

Multimedia Systems
Lecture – 6

By

Dr. Priyambada Subudhi
Assistant Professor
IIIT Sri City

- UTF-16 and UTF-32 suffers from two major problems
  - Unicode uses ASCII for the English language, so A is still 65. However, encoded in 32 bits, the binary representation for the letter A would be 00000000000000000000000001000001. This wastes a lot of valuable space!
  - Many older computers interpret eight zeros in a row (a null) as the end of a string of characters. So these computers wouldn't send any characters that came after eight zeros in a row (they wouldn't send an A if it was represented as 00000000000000000000000001000001).

How UTF -8 encoding works

- The Unicode encoding method **UTF-8** solves these problems of UTF-16 and UTF-32. It works like as follows.
  - Up to character number 128, the regular ASCII value is used (so for example A is 01000001)
  - For any character beyond 128, UTF-8 separates the code into two bytes and adding '110' to the start of first byte to show that it is a beginning byte, and '10' to the start of second byte to show that it follows the first byte.
  - So, for each character beyond number 128, you have two bytes:

    [110xxxxx] [10xxxxxx]   [1100010 1][100001 01]
  - And you just fill in the binary for the number in between:

    [11000101][10000101] (that's the number 325 -> 00101000101)

- This works for the first 2048 characters.

- For characters beyond that, one more '1' is added at the beginning of the first byte and a third byte is also used:

  [1110xxxx] [10xxxxxx] [10xxxxxx]

- This gives you 16 spaces for binary code. In this manner, UTF-8 goes up to four bytes:

  [11110xxx] [10xxxxxx] [10xxxxxx] [10xxxxxx]

- How the Devanagari character क, with code point 2325 can be represented in UTF-8, UTF-16 and UTF-32.

Text File Formats

- Text can be in the form of plain text or rich text.

Plain Text:

- Plain text is a text with no styles attached or embedded with it.
- It is just a plain text i.e. simple text.
- It was created by American Hardware Designers in the 60s 70s. And the files with plain text are saved with .txt or .TXT extension.

Advantages of plain text –

- if you write something in plain text then whenever you open that file on any machine then the end result will be same that is it will show the text as it is on every system. Unlike rich text if you format the text and add the styles to it then it might create problem if the end person doesn't have suitable app for it then it will distort its styles as a result it will leave a bad impression on the user.
- It is fast and flexible as it doesn't uses styles and formatting therefore it is fast and flexible.
- file size of plain text format is less as compared to rich text format.

Disadvantages of plain text –

- In the todays era almost 90% of the word processor uses rich text format.
- It looks very basic i.e. without styles.
- To read the plain text it becomes boring sometimes.

RICH TEXT :

- Rich text is a text with styles attached or embedded with it so whenever you are copying the rich text and pasting to editor which support rich text then it will automatically include all the styles formatting spacing etc.

- Rich text format was created by Microsoft in 80s and they discontinued the development of rich text in 2008. Rich text files are usually save with .rtf or .RTF extension.

Advantages of rich text –

- In almost 90% of word processors supports the rich text format.
- It looks interesting that a text with styles rather then simple text as it include styling and formatting as a result developer learn how to style and format the text.
- To keep the end user connected to the text without getting bored.

Disadvantages of rich text –

- If the rich text file is opened in word processor which do not support rich text then it doesn't looks good to the user as the formatting might not support the word processor.
- file size is large as compared to plain text as it includes styles and formatting information with it
- it is also bit complex to add the styles and formatting as it takes more time compared to plain text

Multimedia Systems
Lecture – 7

By

Dr. Priyambada Subudhi
Assistant Professor
IIIT Sri City

Digital Image Representation

- The most common form to represent natural images and other forms of graphics that are rich in detail is bitmap.
- The term bitmap refers to how a given pattern of bits maps to a specific color.
- They store this information in a grid of points (array of points) or pixels, which has a fixed width and height, and they can store various ranges of colours according to the image type.

A Bitmap image

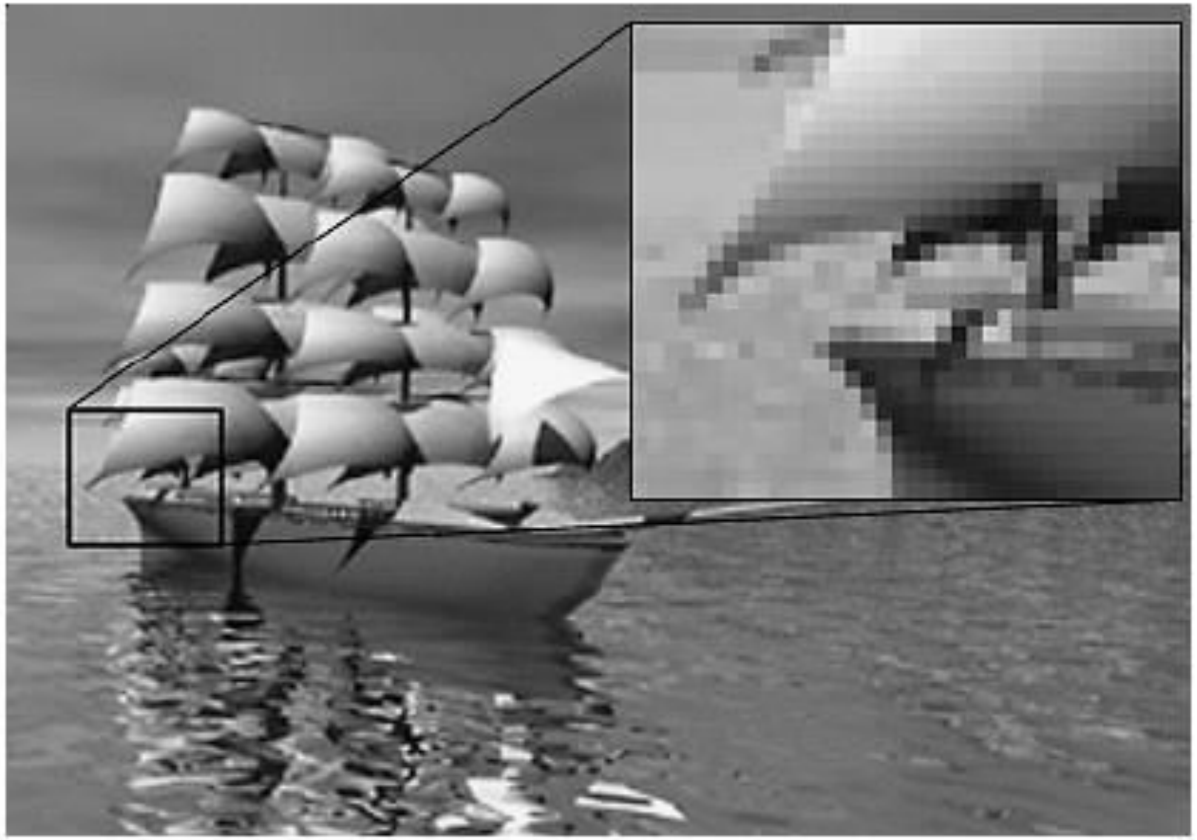


Image Data Type

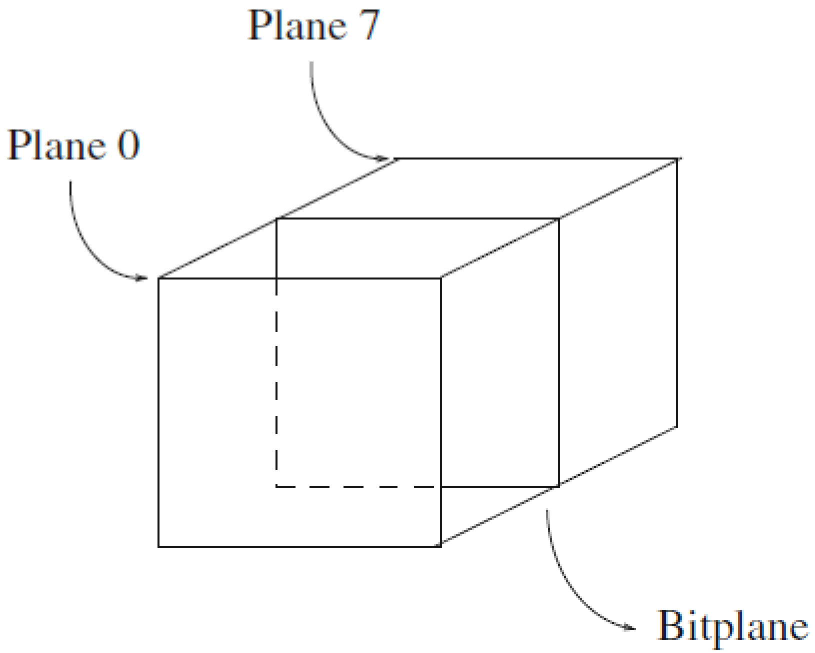
1-Bit Images

- A 1-bit image consists of on and off bits only and thus is the simplest type of image.
- Each pixel is stored as a single bit (0 or 1). Hence, such an image is also referred to as a *binary image*.
- Monochrome 1-bit images can be satisfactory for pictures containing only simple graphics and text.
- Moreover, fax machines use 1-bit data, so in fact 1-bit images are still important even though storage capacities have increased enough to permit the use of imaging that carries more information.

- This is the 1-bit monochrome Lena image of size 640×480.
- A 640×480 monochrome image requires 38.4 kilobytes (kB) of storage.

8-Bit Gray-Level Images

- In an 8-bit image, each pixel has a *gray value* between 0 and 255.
- Each pixel is represented by a single byte—for example, a dark pixel might have a value of 10, and a bright one might be 230.
- We can think of the 8-bit image as a set of 1-bit *bitplanes*, where each plane consists of a 1-bit representation of the image: a bit is turned on if the image pixel has a nonzero value at that bit level.



- Each bitplane can have a value of 0 or 1 at each pixel but, together, all the bitplanes make up a single byte that stores values between 0 and 255
- For the least significant bit, the bit value translates to 0 or 1 in the final numeric sum of the binary number. Positional arithmetic implies that for the next, second bit each 0 or 1 makes a contribution of 0 or 2 to the final sum. The next bits stand for 0 or 4, 0 or 8, and so on, up to 0 or 128 for the most significant bit.
- Each pixel is usually stored as a byte (a value between 0 and 255), so a 640×480 grayscale image requires 300kB of storage (640×480 = 307,200).

Lena gray-level image

24-Bit Color Images

- In a color 24-bit image, each pixel is represented by three bytes, usually representing RGB.
- Since each value is in the range 0–255, this format supports $256 \times 256 \times 256$, or a total of 16,777,216, possible combined colors.
- However, such flexibility does result in a storage penalty: a 640×480 24-bit color image would require 921.6kB of storage without any compression.

Resolution

- Resolution is a measure of how finely a device displays graphics with pixels. It is used by printers, scanners, monitors (TV, computer), mobile devices and cameras.
- There are two types ways of measuring resolution:
  - The amount of (dpi) dots per inch. Printers and scanners work with higher resolutions than computer monitors. Current desktop printers can support 300dpi +, flatbed scanners from 100-3600dpi+. In comparison computer monitors support 72-130 dpi.
  - ppi (Pixels Per Inch) is a term also used to define the resolution for bitmaps. Bitmap images are resolution-dependent as they contain a fixed number of pixels.
  - As a result, they can lose detail and appear jagged if they are scaled to high magnifications on screen or if they are printed at a lower resolution than they were created for.

Example of a Bitmap Image at Different Levels of Magnification

- Bitmap images contain a fixed number of pixels, usually measured in pixels per inch (ppi). An image with a high resolution contains more, and therefore smaller, pixels than an image of the same printed dimensions with a low resolution.

- *Example:* A one inch by one inch image with a resolution of 72 ppi contains a total of 5184 pixels (72 pixels wide × 72 pixels high = 5184). The same one inch by one inch image with a resolution of 300 ppi would contain a total of 90,000 pixels.

Multimedia Systems
Lecture – 8

By

Dr. Priyambada Subudhi
Assistant Professor
IIIT Sri City

Assessment Plan

Sl. No.	Mode of Assessment	% of Marks
1	Mid Semester	20
2	End Semester	30
3	Scheduled Quiz	20
4	Surprised Quiz	10
5	Project	20

8-bit color images

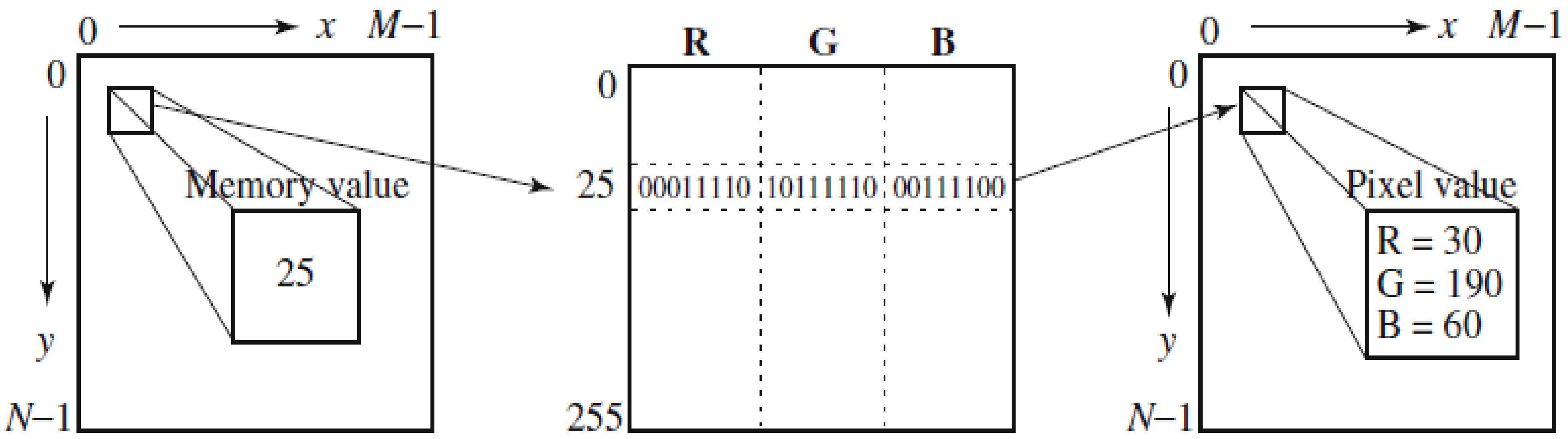
- Many systems can utilize color information stored with only 8 bits of information (the so-called 256 colors) in producing a screen image.

- Such image files use the concept of a lookup table to store color information.
  - Basically, the image stores not color, but instead just a set of bytes, each of which is actually an index into a table with 3-byte values that specify the color for a pixel with that lookup table index.

Color Lookup Tables

- The idea used in 8-bit color images is to store only the index, or code value, for each pixel.
- Then, if a pixel stores, say, the value 25, the meaning is to go to row 25 in a color lookup table (LUT).
- For an 8-bit image, the image file can store in the file header information just what 8-bit values for R, G, and B correspond to each index.
- The LUT is often called a *palette.*

Color LUT for 8-bit color images

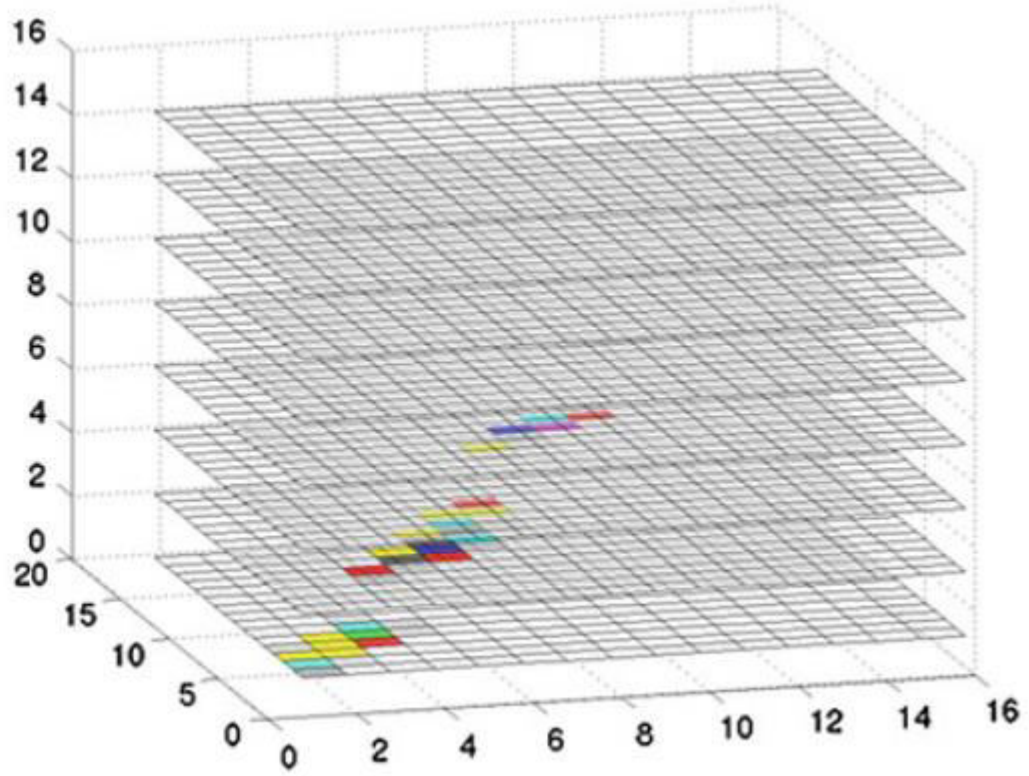


Color Histogram

- It makes sense to carefully choose just which colors to represent best in the image.
  - if an image is mostly red sunset, it is reasonable to represent red with precision and store only a few greens.
- Suppose all the colors in a 24-bit image were collected in a $256 \times 256 \times 256$ set of cells, along with the count of how many pixels belong to each of these colors stored in that cell.
  - For example, if exactly 23 pixels have RGB values (45, 200, 91) then store the value 23 in a three-dimensional array, at the element indexed by the index values [45, 200, 91]
- This data structure is called a *color histogram*.

3D scatterplot of RGB colors in forestfire image

- The histogram has 16 × 16 × 16 bins and shows the count in each bin in terms of intensity and pseudocolor.
- We can see a few important clusters of color information, corresponding to the reds, yellows, greens, and so on, of the forestfire image.
- Basically, large populations in 3D histogram bins can be subjected to a split-andmerge algorithm to determine the “best” 256 colors



Example of an 8-bit color image

Note the great savings in space for 8-bit images over 24-bit ones: a 640 × 480 8-bit color image requires only 300kB of storage, compared to 921.6kB for a color image.

24-bit color image

8-bit color image

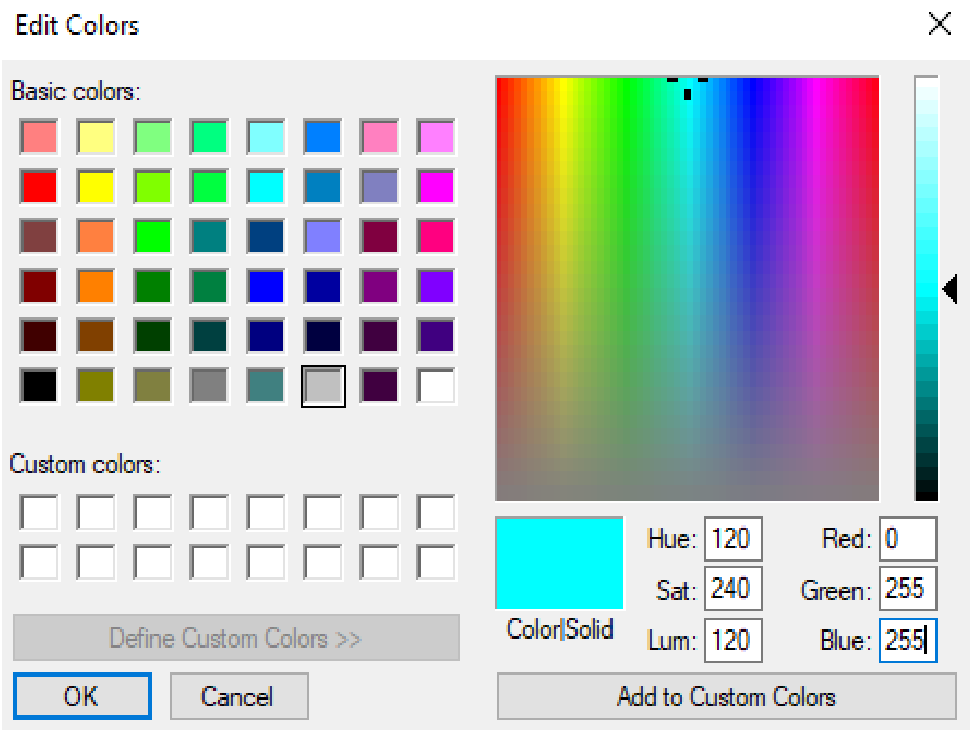
How to Transform to 8 bit RGB

How to Select the Best 256 LUT RGB Entries Without Constructing a Color Histogram

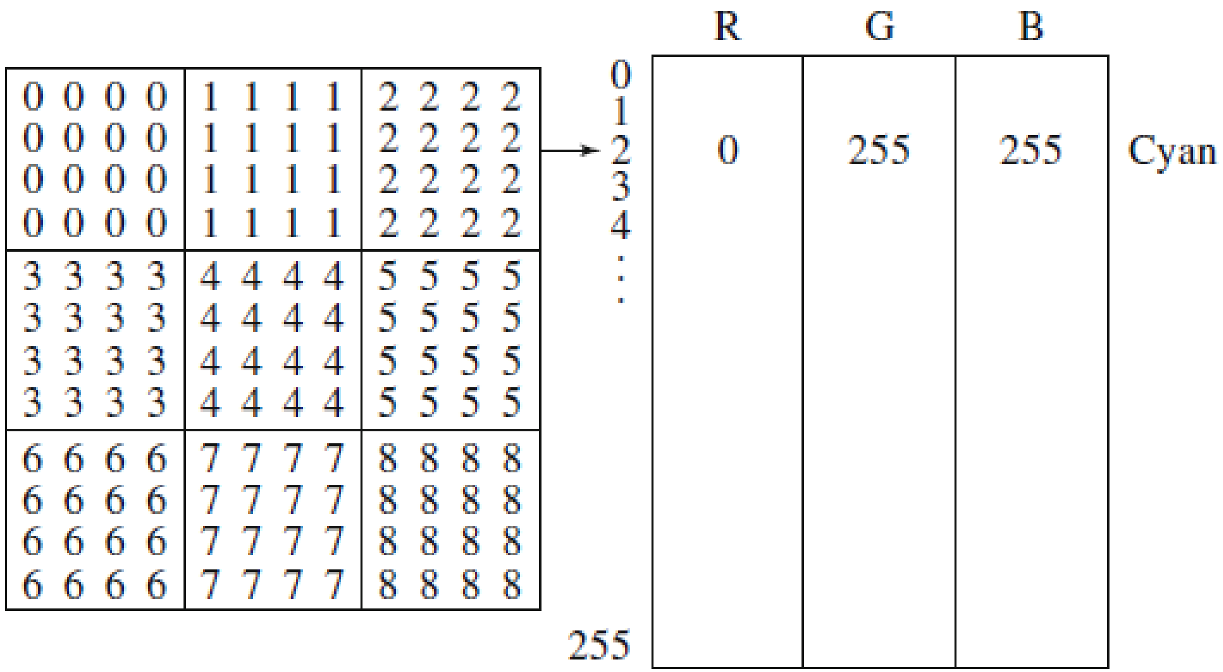
- Transform 3 bytes RGB to 8bit per pixel for 256 selective LUT table index by selecting RGB color separately.
- Sort R byte in the original image to select 8 (3bits) most popular R intensities.
  - Those 8 most popular R byte codes to LUT.
- Sort G byte to select 8 (3bits) most popular G intensities
  - Those 8 most popular G byte codes to LUT
- Sort B byte to select 4 (2bits) most popular B intensities.
  - Those 4 B byte codes to LUT
- All the combinations of 8R x 8G x 4B becomes 256 LUT RGB entries.
- Transformed 3+3+2 bits per pixel is an index to transformed LUT.
- Each pixel value in the original RGB is transformed to index value of the nearest RGB entries in LUT by comparing the original to RGB LUT entries.

Color Picker

- A *color picker* consists of an array of fairly large blocks of color (or a semicontinuous range of colors) such that a mouse click will select the color indicated.
- In reality, a color picker displays the palette colors associated with index values from 0 to 255.
- If the user selects the color block with index value 2, then the color meant is cyan, with RGB values (0, 255, 255).



Color picker for 8-bit color: each block of the color picker corresponds to one row of the color LUT



Multimedia Systems
Lecture – 10

By

Dr. Priyambada Subudhi
Assistant Professor
IIIT Sri City

JPEG

- The most important current standard for image compression is JPEG.
- This standard was created by a working group of the International Organization for Standardization (ISO) that was informally called the *Joint Photographic Experts Group* and is therefore so named.
- The human vision system has some specific limitations, which JPEG takes advantage of to achieve high rates of compression.
- The eye–brain system cannot see extremely fine detail.
- If many changes occur within a few pixels, we refer to that image segment as having *high spatial frequency* —that is, a great deal of change in (x, y) space.

- Therefore, color information in JPEG is decimated (partially dropped, or averaged) and then small blocks of an image are represented in the spatial frequency domain (u, v), rather than in (x, y).
- That is, the speed of changes in x and y is evaluated, from low to high, and a new “image” is formed by grouping the coefficients or weights of these speeds.
- Weights that correspond to slow changes are then favored, using a simple trick: values are divided by some large integer and truncated. In this way, small values are zeroed out.
- Since we effectively throw away a lot of information by the division and truncation step, this compression scheme is “*lossy*”
- JPEG allows the user to set a desired *level of quality*, or *compression ratio* (input divided by output).

JPEG image with low quality specified by user.

This image is having a quality factor $Q = 10$. (The usual default quality factor is $Q = 75$). This image is a mere 1.5% of the original size. In comparison, a JPEG image with $Q = 75$ yields an image size 5.6% of the original, whereas a GIF version of this image compresses down to 23.0% of the uncompressed image size.

PNG

- **PNG format**: standing for ***Portable Network Graphics*** — meant to supersede the GIF standard, and extends it in important ways.
- Special features of PNG files include:
  - support for up to 16 bits per pixel in each color channel, i.e., 48-bit color—a large increase.
  - Files may contain gamma-correction information for correct display of color images, as well as alpha-channel information for such uses as control of transparency.
  - The display progressively displays pixels in a 2-dimensional fashion by showing a few pixels at a time over seven passes through each 8 x 8 block of an image.
  - It supports both lossless and lossy compression with performance better than GIF. PNG is widely supported by various web browsers and imaging software.

TIFF

- **TIFF**: stands for *Tagged Image File Format* is another popular image file format. Developed by the *Aldus Corporation* in the 1980s, it was later supported by *Microsoft*.
- The support for attachment of additional information (referred to as "tags") provides a great deal of flexibility.
- The most important tag is a *format signifier*: what type of compression etc. is in use in the stored image.
- TIFF can store many different types of image: 1-bit, grayscale, 8-bit color, 24-bit RGB, etc.
- TIFF was originally a lossless format but now a JPEG tag allows one to opt for JPEG compression.

EXIF

- **EXIF** (*Exchange Image File*) is an image format for digital cameras.
- It enables the recording of image metadata (exposure, light source/flash, white balance, type of scene, etc.) for the standardization of image exchange.
- A variety of tags (many more than in TIFF) is available to facilitate higher quality printing, since information about the camera and picture-taking conditions can be stored and used, e.g., by printers for possible color-correction algorithms.
- The EXIF format is incorporated in the JPEG software in most digital cameras.

PS and PDF

- ***Postscript*** is an important language for typesetting, and many high-end printers have a Postscript interpreter built into them.
- Postscript is a vector-based picture language, rather than pixel based: page element definitions are essentially in terms of vectors.
- Postscript includes text as well as vector/structured graphics.
- Bit-mapped images can be included in output files.
- *Encapsulated Postscript files (.EPS)* add some additional information for inclusion of Postscript files in another document.
- Postscript page description language itself does not provide compression; in fact, Postscript files are just stored as ASCII.

- For files containing images, PDF may achieve higher compression ratios by using separate JPEG compression for the image content.
- Another text + figures language has superseded or at least paralleled Postscript: Adobe Systems Inc. includes LZW compression in its *Portable Document Format (**PDF**)* file format.
- PDF files that do not include images have about the same compression ratio, 2:1 or 3:1, as do files compressed with other LZW-based compression tools.
- A useful feature of the Adobe Acrobat PDF reader is that it can be configured to read documents structured as linked elements, with clickable content and handy summary tree-structured link diagrams provided.

Some Other Image Formats

- **Microsoft Windows: WMF (***Windows MetaFile***)**: the native vector file format for the Microsoft Windows operating environment:
- 1. Consist of a collection of GDI (Graphics Device Interface) function calls, also native to the Windows environment.
- 2. When a WMF file is "played" (typically using the Windows PlayMetaFile() function) the described graphics is rendered.
- 3. WMF files are ostensibly device-independent and are unlimited in size.

- **Microsoft Windows: BMP (*Bitmap image files*)**: the major system standard graphics file format for Microsoft Windows, recognized by many programs. Watch it!: there are many sub-variants within the BMP standard.

- **Netpbm Format**: PPM (Portable PixMap), PGM (Portable GrayMap), and PBM (Portable BitMap) belong to a family of open source Netpbm formats. These formats are mostly common in the linux/unix environments.

PTM

- PTM (*Polynomial Texture Mapping*) is a technique for storing a representation of a camera scene that contains information about a set of images taken under a set of lights that each have the same spectrum (say, a xenon flash), but with each light placed at a different direction from the scene. PTM was invented at Hewlett-Packard.

- **a** 50 input images for PTM: lights individually from 50 different directions ***e****i* , *i* = 1 . . . 50; **b** interpolated image under new light ***e***

Multimedia Systems Lecture – 11

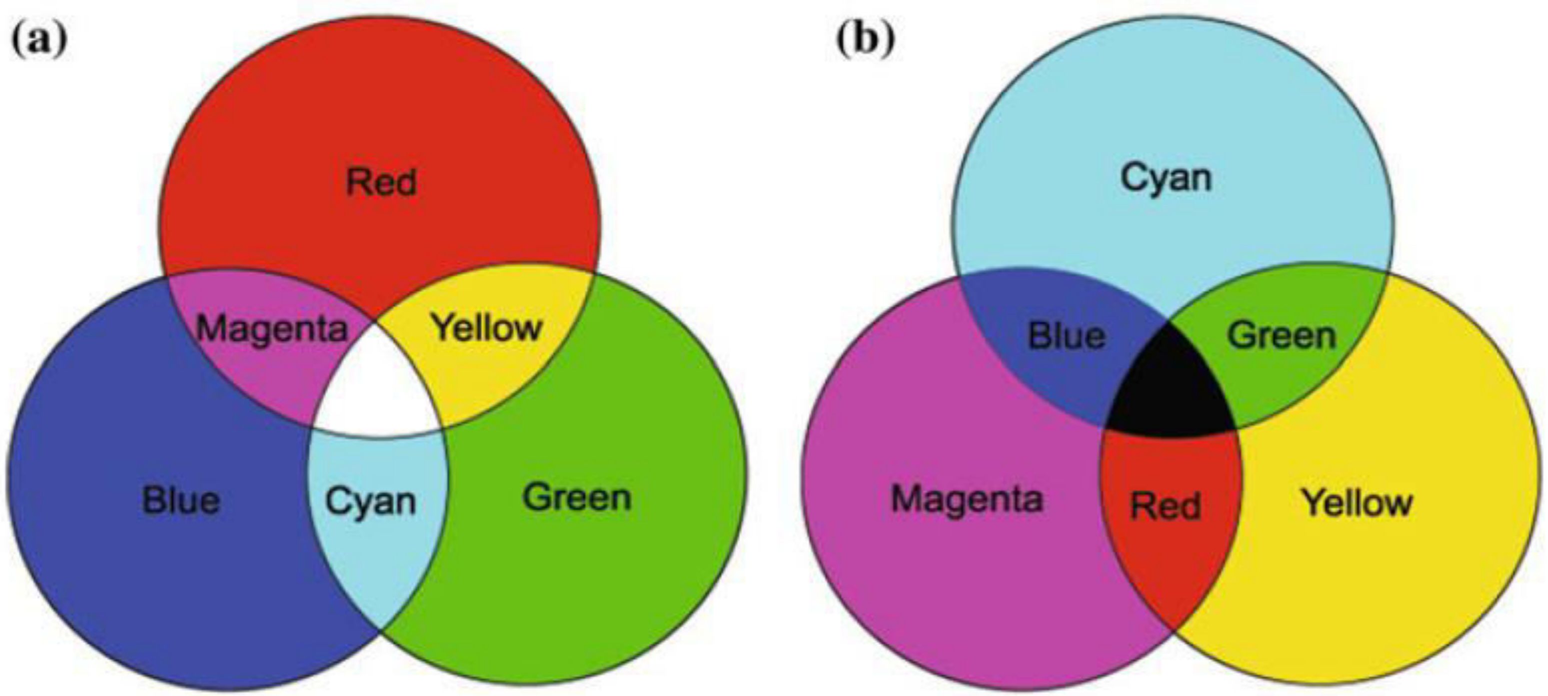
By

Dr. Priyambada Subudhi
Assistant Professor
IIIT Sri City

Color Model

- A color model is an orderly system for creating a whole range of colors from a small set of ***primary colors*** *(are the set of colors that can be combined to make an useful range of colors)*.
- *Color Gamut*: Set of all colors that we can produce from the primary colors.
- There are two types of color models.
  - Additive color model (e.g. RGB color model)
  - Subtractive color model (e.g. CMY color model)
- Additive color models use light to display color while subtractive models use printing inks.
- Colors perceived in additive models are the result of transmitted light while the colors perceived in subtractive models are the result of reflected light.

- There are several established color models used in computer graphics, but the two most common are the *RGB model* (***Red-Green-Blue***) for computer display and the *CMYK model* (***Cyan-Magenta-Yellow-Black***) for printing.

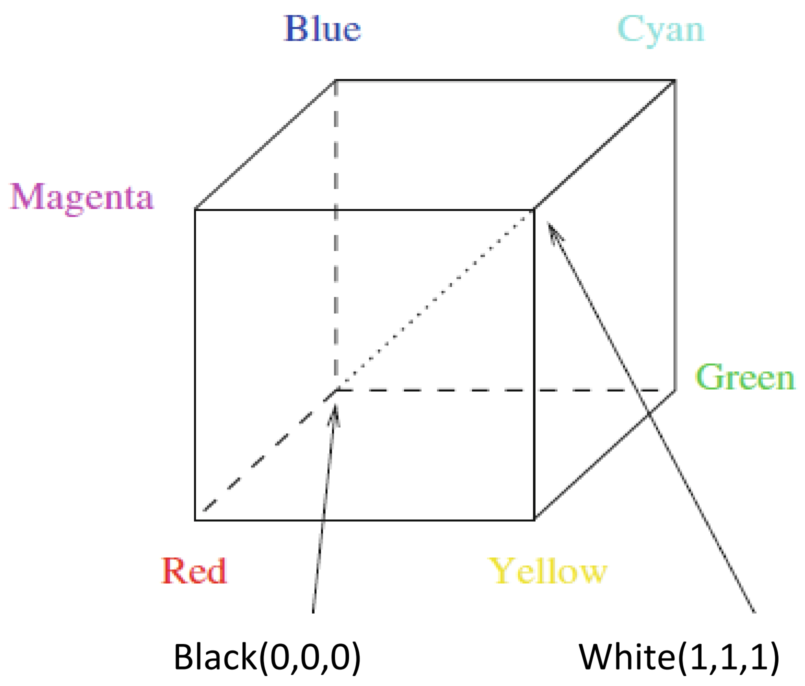


Additive and subtractive color. **a** RGB is used to specify additive color. **b** CMY is usedto specify subtractive color

RGB Model

- The red, green, and blue (RGB) color space is widely used throughout computer graphics.
- Unit Cube defined on R, G & B axes.
- The Origin **(0,0,0)** represents black and the diagonally opposite vertex **(1,1,1)** is White.
- Vertices of the cube on the axes represent primary colors, and the remaining vertices are the complementary color points for each of the primary colors.
- Shades of gray are represented along the main diagonal.

RGB color Cube



- Each color point within the unit cube can be represented as w weighted vector sum of the primary colors, using unit vectors **R**,**G** and **B**.

$$C(\lambda) = R\mathbf{R} + G\mathbf{G} + B\mathbf{B}$$

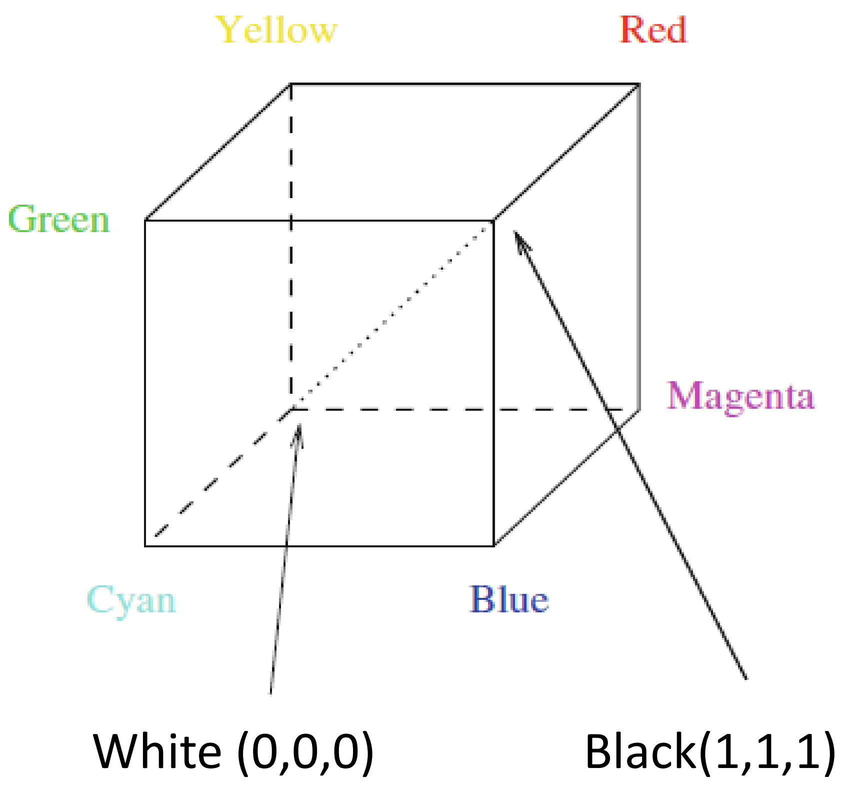
Where R, G and B are the assigned values in the range from 0 to 1.0

- For example, the magenta vertex is obtained by adding the maximum red and blue values to produce : (1,0,1)

CMY and CMYK model

- Subtractive Color Model.
- Stands for cyan-magenta-yellow.
- Used for hardcopy devices (ex. Printers).
- A printed color that looks red absorbs the other two components G and B and reflects R.
- Thus the C-M-Y coordinates are just the complements of the R-G-B coordinates.

CMY color Cube



- In additive color models such as RGB, white is the “additive” combination of all primary colored lights, while black is the absence of light.
- In the CMYK model, it is the opposite: white is the natural color of the paper or other background, while black results from a full combination of colored inks.
- **Transformation from RGB to CMY and CMY to RGB**

$$\begin{bmatrix} C \\ M \\ Y \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} R \\ G \\ B \end{bmatrix} \qquad \begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} C \\ M \\ Y \end{bmatrix}$$

- **[*Complementary colors*:** *Pairs of colors which, when combined in the right proportions, produce white. Example, in the RGB model: red & cyan , green & magenta , blue & yellow.]*

CMYK Model

- Although cyan, magenta and yellow inks might be expected be sufficient for color printing, most actual color printing uses black ink in addition.
- This is partly because a mixture of the first three inks may not yield a black that is neutral enough, or dark enough, but also because the use of black spares the use of the more expensive colored inks, and also reduces the total amount of ink used, thus speeding drying times.
- K used instead of equal amounts of CMY

- Truly "black" black ink is in fact cheaper than mixing colored inks to make black, so a simple approach to producing sharper printer colors is to calculate that part of the three-color mix that would be black, remove it from the color proportions, and add it back as real black. This is called "undercolor removal."
- With *K* representing the amount of black, the new specification of inks is thus

$$K \equiv \min\{C, M, Y\}$$

$$\begin{bmatrix} C \\ M \\ Y \end{bmatrix} \Rightarrow \begin{bmatrix} C - K \\ M - K \\ Y - K \end{bmatrix}$$

Multimedia Systems
Lecture – 12

By

Dr. Priyambada Subudhi

Assistant Professor

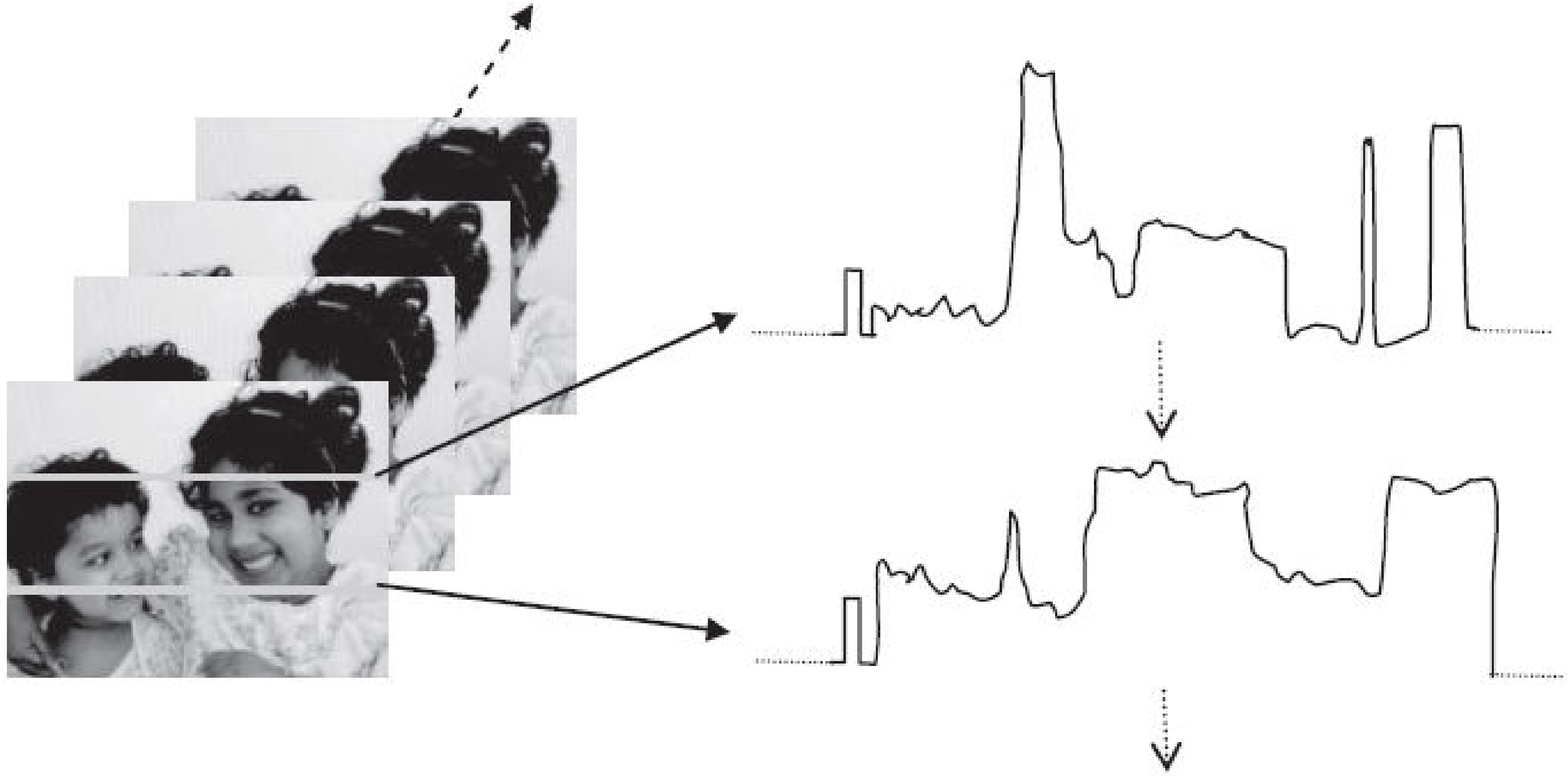
IIIT Sri City

Video

- Video, whether analog or digital, is represented by a sequence of discrete images shown in quick succession. Each image in the video is called a frame, which is represented as a matrix of pixels defined by a width, height, and pixel depth.
- In addition, two important properties govern video representation: frame rate and scanning format.
- The rate at which the images are shown is the frame rate.
- If the frame rate is too slow, the human eye perceives an unevenness of motion called flicker.

- Although digital video can be considered a three-dimensional signal—a 2D image changing over time—analog video is converted to a 1D signal of scan lines.
- This scan line conversion was introduced to make analog television broadcast technology work, and is central to the manner in which televisions (and all other cathode-ray tubes) display images.
- The electron gun(s) in a television project electrons on the phosphor screen from left to right in a scan line manner and from top to bottom successively for each frame.
- The phosphor screen glows at each location on a scan line creating a color at all positions on the line.
- Scanning formats, which is an outcome of the analog technology, can be represented as interlaced or progressive.

- *Left: Video is represented as a sequence of images. Right: Analog video of one frame scanned as a 1D signal. Each scan line is scanned from left to right as an analog signal separated by horizontal syncs. Two scan lines are shown; each begins with a horizontal sync and traces through the intensity variation on that scan line.*

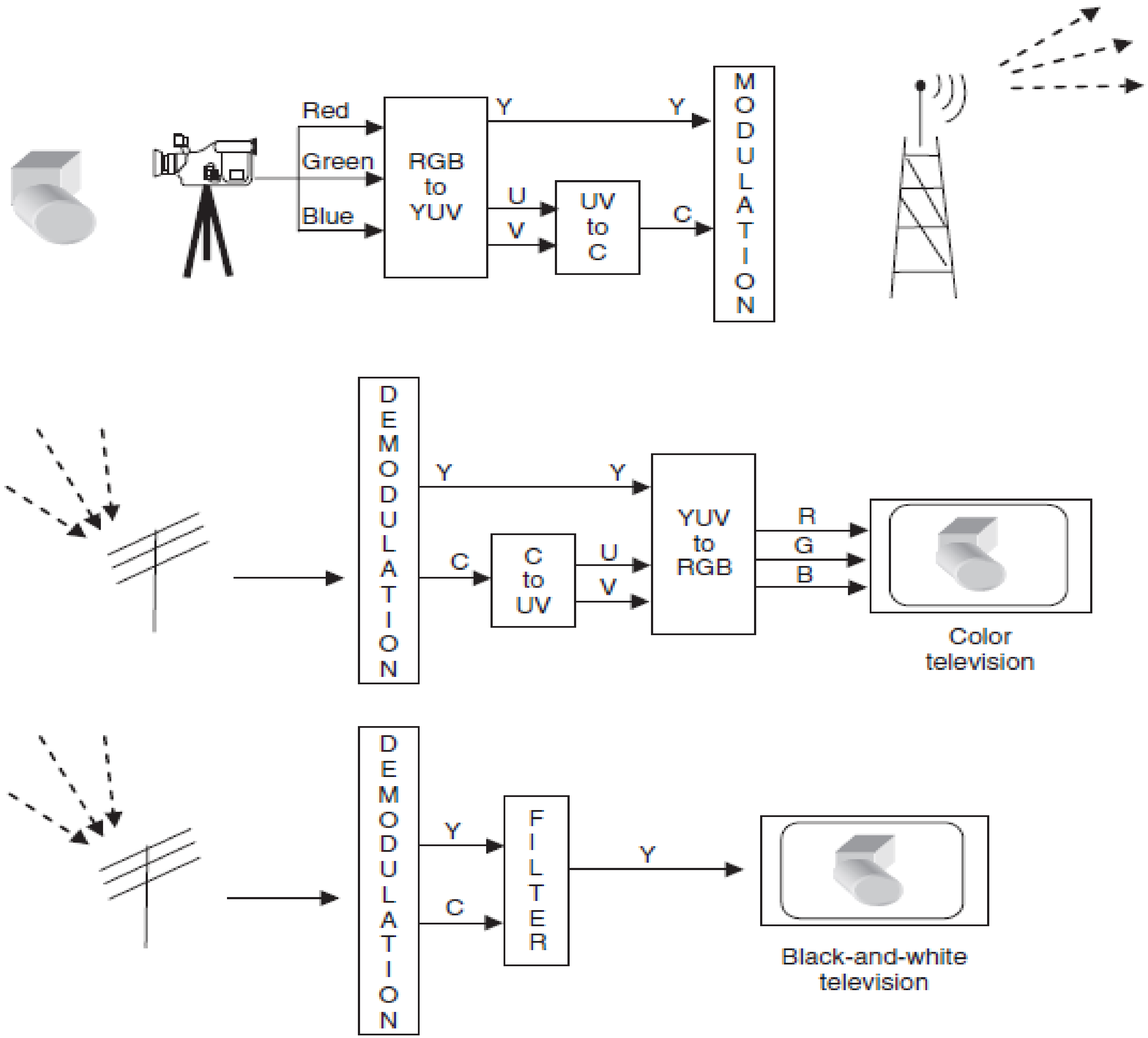


Digital display technologies display media in a digital format. Digital video display on these devices, such as LCD or plasma, does not require the scanning mechanism described previously. However, when the technology for digital video started to evolve, the television instruments were still rendering analog signals only. As a result, the digital video standards have their representations and formats closely tied to analog TV standards.

Analog Video and Television

- Analog video signal used in broadcast is scanned as a one-dimensional signal in time, where the spatiotemporal information is ordered as a function of time according to a predefined scanning convention.
- This 1D signal captures the time-varying image intensity information only along scanned lines.
- Television requires this analog scanned information to be broadcast from a broadcast station to all users.
- The standardization process implemented in the broadcast of analog video for television mandated a few requirements, which were necessary for making television transmission viable: YUV color space conversion and interlaced scanning.

Television works by sending scan line information in interlaced YUV format.



Conversion to YUV

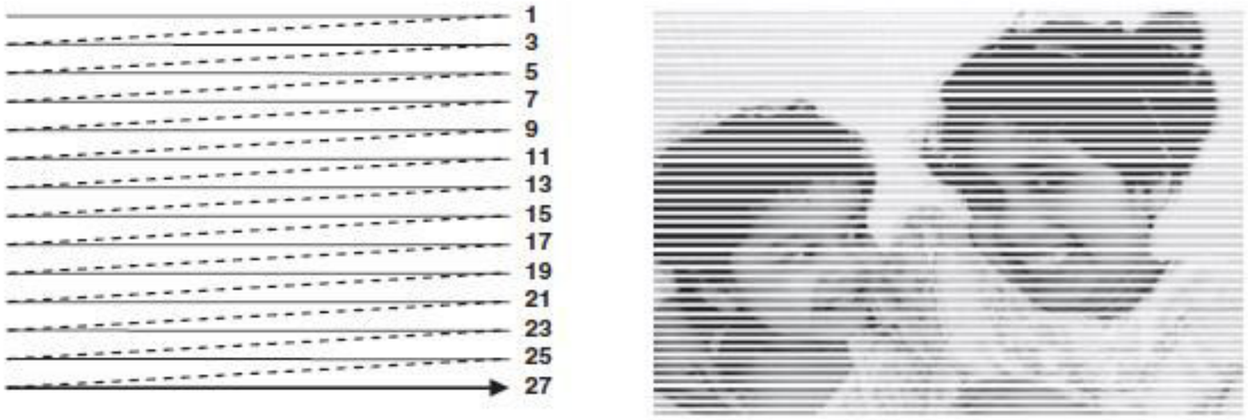
- Video frames, like images, are represented using a color format, which is normally RGB. This RGB color space is used by cathode-ray tube–based display devices, such as the television, to display and render the video signal.
- For transmission purposes, however, the RGB signal is transformed into a YUV signal. The YUV color space aims to decouple the intensity information (Y or luminance) from the color information (UV or chrominance).
- The separation was intended to reduce the transmission bandwidth and is based on experiments with the human visual system, which suggests that humans are more tolerant to color distortions than to intensity distortions.

Analog Video Scanning

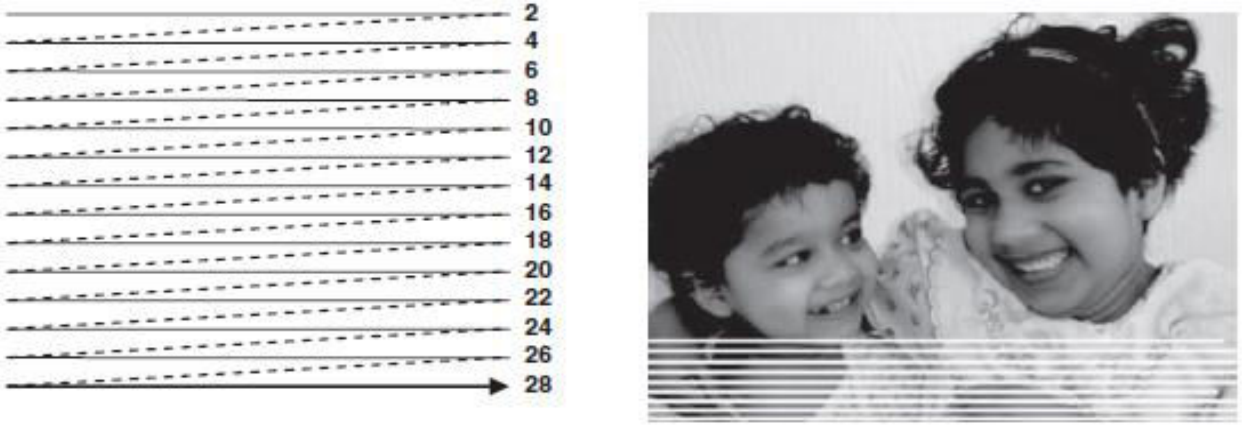
- Video is scanned as a 1D signal, where each raster line is interspaced with horizontal and vertical syncs.
- For horizontal synchronization in analog video, a small voltage offset from zero is used to indicate black and another value, such as zero, to indicate the start of a line.
- Vertical synchronization is carried out by the cycles in the power outlet (60 Hz for NTSC, 50 Hz for PAL). Every 1/60th of a second, the electron gun is reset by the vertical sync to draw the beginning of the next frame.

- In TV and in some monitors and multimedia standards, another system, *interlaced* scanning, is used.
- Here, the odd-numbered lines are traced first, then the even-numbered lines. This results in “odd” and “even” *fields*—two fields make up one frame.
- But the resulting video drawn by interlaced scanning techniques might be unacceptable and has occasional flicker and artifacts. This is caused because the video is captured at different moments in time as two field and, hence, interlaced video frames exhibit motion artifacts when both fields are combined and displayed at the same moment.
- Video is of better quality when it is captured progressively and drawn progressively, which eliminates the occasional flicker.

- *Interlaced scanning. The top figure shows the upper "odd" field consisting of odd-numbered lines. The bottom shows a lower "even" field interspersed with the odd field. Both fields are shown in succession to meet the required frame rate.*

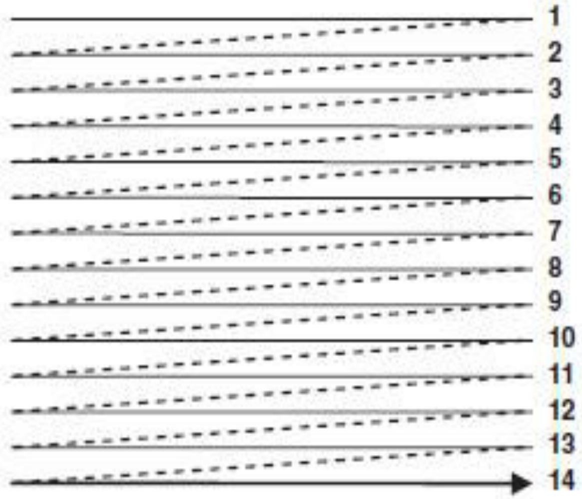


Upper field



Lower field

Progressive scanning. All the scan lines are drawn in succession, unlike in the interlaced case.



Analog Video Standards

NTSC Video

- ***NTSC***, named for the ***National Television System Committee***, is the analog television system that is mostly used in most of North America and Japan.
- It uses a familiar 4:3 *aspect ratio* (i.e., the ratio of picture width to height) and 525 scan lines per frame at 30 fps.
- NTSC video is an analog signal with no fixed horizontal resolution. Therefore, we must decide how many times to sample the signal for display. Each sample corresponds to one pixel output. A *pixel clock* divides each horizontal line of video into samples. The higher the frequency of the pixel clock, the more samples per line.

PAL Video

- *PAL (Phase Alternating Line)* is a TV standard originally invented by German scientists.
- It uses 625 scan lines per frame, at 25 frames per second (or 40 msec / frame), with a 4 : 3 aspect ratio and interlaced fields.
- This important standard is widely used in Western Europe, China, India and many other parts of the world.
- PAL uses the YUV color model with an 8 MHz channel, allocating a bandwidth of 5.5 MHz to Y and 1.8 MHz each to U and V.
- To improve picture quality, chroma signals have alternate signs (e.g., +U and — U) in successive scan lines; hence the name "Phase Alternating Line.
- This facilitates the use of a (line - rate) comb filter at the receiver — the signals in consecutive lines are averaged so as to cancel the chroma signals (which always carry opposite signs) for separating Y and C and obtain high - quality Y signals.

SECAM Video

- SECAM, which was invented by the French, is the third major broadcast TV standard.
- SECAM stands for Systeme Electronique Couleur Avec Memorie.
- SECAM also uses 625 scan lines per frame, at 25 frames per second, with a 4:3 aspect ratio and interlaced fields.
- SECAM and PAL are similar, differing slightly in their color coding scheme.
- In SECAM, U and V signals are modulated using separate color subcarriers at 4.25 MHz and 4.41 MHz, respectively. They are sent in alternate lines - that is, only one of the U or V signals will be sent on each scan line.

Multimedia Systems
Lecture – 13

By

Dr. Priyambada Subudhi
Assistant Professor
IIIT Sri City

Types of Video Signals

Composite Video

- Composite video is also called baseband video or RCA video. It is the analog waveform that conveys the image data in the conventional NTSC, PAL and SECAM television signal.
- Composite video contains both chrominance (color) and luminance (brightness) information, along with synchronization and blanking pulses, all together in a single signal.
- This was done to reduce bandwidth and achieve real-time transmission.
- However, in composite video, interference between the chrominance and luminance information is inevitable and tends to worsen when the signal is weak.
- This is why fluctuating colors, false colors, and intensity variations are seen when a distant NTSC television station sends signals that are weak and not properly captured at home with old-fashioned “rabbit ears,” or outdoor “aerial” antennae.

S-Video

- S-Video (*Super-Video*, sometimes referred to as *Y/C Video*) is a video signal transmission in which the luminance signal and the chrominance signal are transmitted separately to achieve superior picture clarity.
- The luminance signal (*Y*) carries brightness information, and the chrominance signal (*C*) carries color information.
- Here, the chrominance signal (*C*) is formed by combining the two chrominance signals *U* and V into one signal along with their respective synchronization data, so at display time, the *C* signal can be separated into *U* and *V* signals.
- Separating the *Y* and *C* channels and sending them separately reduces problems caused by interference between the luminance and chrominance signals and yields a superior visual quality.

Component Video

- Component video strives to go a step further than S-Video by keeping all three *Y, U, V* (or equivalent) components separate.
- Consequently, the bandwidth required to broadcast component video is more than the composite or S-Video and, correspondingly, so is the visual quality.
- The separation of these components prevents artifacts due to intersignal interference.

Connectors for typical analog display interfaces. From left to right:, Composite video, S-video, and Component video

Digital Video

The advantages of digital representation for video are many. It permits

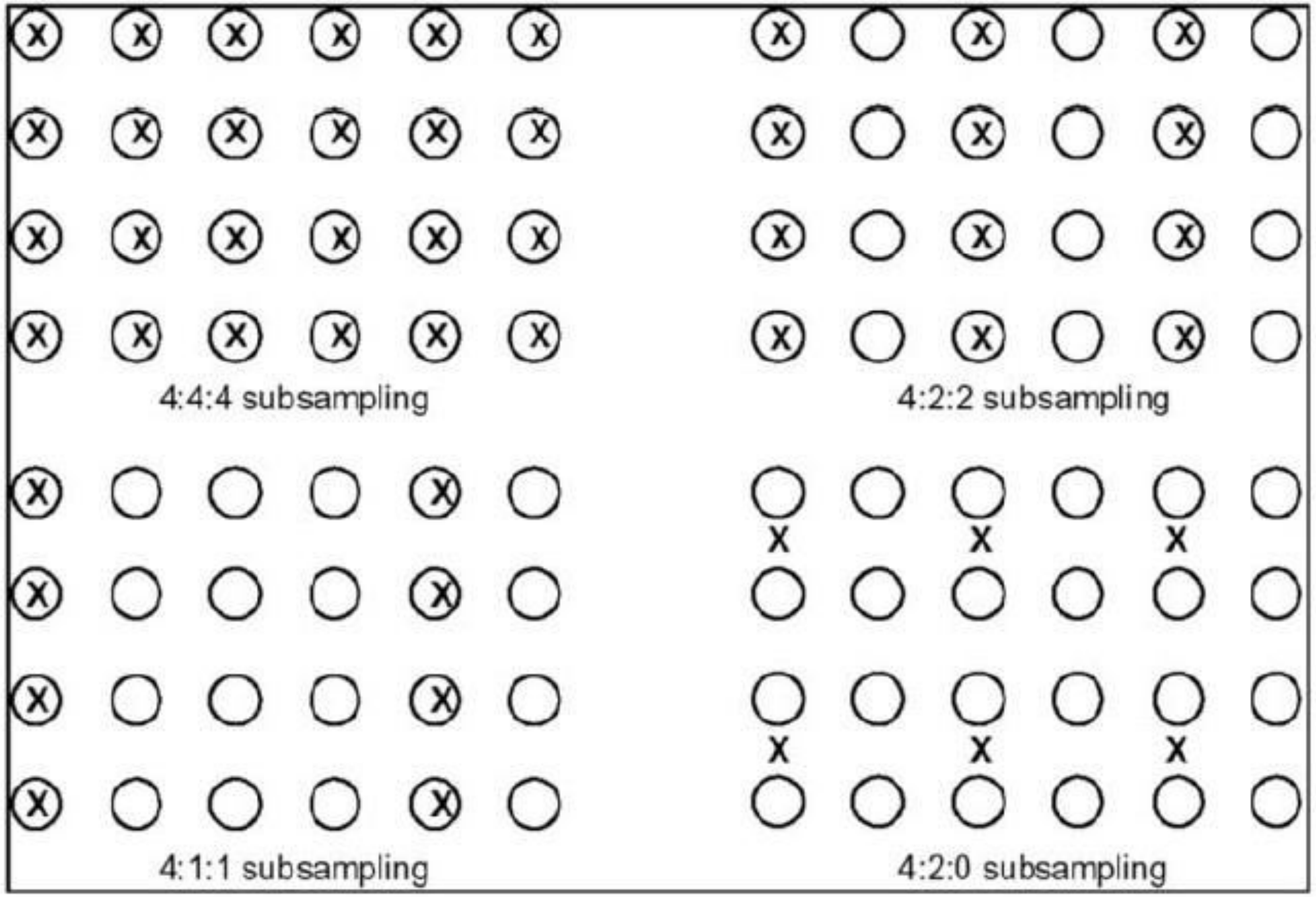
- Storing video on digital devices or in memory, ready to be processed (noise removal, cut and paste, and so on) and integrated into various multimedia applications.
- Direct access, which makes nonlinear video editing simple.
- Repeated recording without degradation of image quality.
- Ease of encryption and better tolerance to channel noise.

YUV Subsampling Schemes

- Video signals captured by digital cameras are represented in the RGB color space, which is also used to render video frames on a display device.
- However, for transmission and other intermediary processing, the YUV space is commonly used.
- The YUV space separates the color and luminance information.
- The color information (UV) is then further subsampled to gain more bandwidth.
- In analog video, subsampling is achieved by allocating half as much bandwidth to chrominance as to luminance.
- In digital video, subsampling can be done by reducing the number of bits used for the color channels on average.

- Depending on the way subsampling is done, a variety of subsampling ratios can be achieved.
- The circles represent pixel information.
- Potentially, we could store 1 byte each for Y, U, and V components, resulting in 24 bits per pixel.
- In subsampling, the luminance component Y is left untouched—that is, 1 byte is reserved for the luminance data per pixel.
- An X at a pixel position suggests that we also store the chrominance components for this position.

YUV subsampling schemes used in video



- In the 4:4:4 scheme, each pixel has luminance (8 bits) and chrominance (8 bits for U and 8 bits for V), resulting in 24 bits per pixel.
- In the 4:2:2 subsampling scheme, chrominance information is stored for every other pixel bringing the equivalent bits per pixel down to 16.
- In the 4:1:1 subsampling scheme, chrominance is stored every fourth pixel in a row.
- Whereas in the 4:2:0 scheme, the average of the U values for a 2×2 pixel area is stored, and similarly for the V values.
- Since there is only 1 U and 1 V sample for every four luminance samples, the equivalent bits per pixel is brought down to 12 bits per pixel.

Multimedia Systems Lecture – 14

By

Dr. Priyambada Subudhi

Assistant Professor

IIIT Sri City

CCIR and ITU-R Standards for Digital Video

- The CCIR is the *Consultative Committee for International Radio*. One of the most important standards it has produced is *CCIR-601* for component digital video.
- This standard has since become standard ITU-R Rec. 601, an international standard for professional video applications.
- It is adopted by several digital video formats, including the popular DV video.
- The NTSC version has 525 scan lines, each having 858 pixels. Because the NTSC version uses 4:2:2, each pixel can be represented with two bytes (8 bits for Y and 8 bits alternating between U and V). The Rec. 601 (NTSC) data rate is thus approximately 216Mbps.

- The CIF format (Common Interchange Format) was established for a progressive digital broadcast television. It consists of VHS quality resolutions whose width and height are divisible by 8—a requirement for digital encoding algorithms.
- The Quarter Common Interchange Format (QCIF) was established for digital videoconferencing over ISDN lines.
- CIF is a compromise between NTSC and PAL, in that it adopts the NTSC frame rate and half the number of active lines in PAL.
- When played on existing TV sets, NTSC TV will first need to convert the number of lines, whereas PAL TV will require frame rate conversion.

ITU-R digital video specifications

	Rec. 601 525/60 NTSC	Rec. 601 625/50 PAL/SECAM	CIF	QCIF
Luminance resolution	720×480	720×576	352×288	176×144
Chrominance resolution	360×480	360×576	176×144	88×72
Color subsampling	4:2:2	4:2:2	4:2:0	4:2:0
Aspect ratio	4:3	4:3	4:3	4:3
Fields/sec	60	50	30	30
Interlaced	Yes	Yes	No	No

High-Definition TV (HDTV)

- The usual NTSC analog TV signal in the United States has 525 scan lines, with 480 actually visible. The usual TV has an effective picture resolution of about 210,000 pixels.
- Today, consumers are accustomed to better resolutions such as 1024×768 and even higher, which are now commonly supported by most graphics hardware that come with computers.
- A class of digital television called HDTV supports a higher resolution display format along with surround sound.
- The visual formats used in HDTV are as follows:
  - 720*p*—1280×720 pixels progressive
  - 1080*i*—1920×1080 pixels interlaced
  - 1080*p*—1920×1080 pixels progressive

- They use the MPEG2-based video compression format with a 17 Mbps bandwidth.
- Although HDTV signals can be stored and transmitted effectively using MPEG-2 technology, a lot of bandwidth is required to transmit numerous channels.
- The aspect ratio of HDTV is 16:9 (1.78:1), which is closer to the ratios used in theatrical movies, typically 1.85:1 or 2.35:1.
- The increased resolution provides for a clearer, more detailed picture. In addition, progressive scan and higher frame rates result in a picture with less flicker and better rendering of fast motion.

Ultra High Definition TV (UHDTV)

- UHDTV is a new development—a new generation of HDTV.
- The standards announced in 2012 support 4K UHDTV: 2160P (3,840×2,160, progressive scan) and 8K UHDTV: 4320P (7,680 × 4,320, progressive scan).
- The aspect ratio is 16:9. The bit-depth can be up to 12 bits, and the chroma subsampling can be 4:2:0 or 4:2:2.
- The supported frame rate has been gradually increased to 120 fps.
- The UHDTV will provide superior picture quality, comparable to IMAX movies, but it will require a much higher bandwidth and/or bitrate.

Digital Display Interfaces

- Given the rise of digital video processing and the monitors that directly accept digital video signals, there is a great demand toward video display interfaces that transmit digital video signals.

- The most widely used digital video interfaces include
  - Digital Visual Interface (DVI)
  - High- Definition Multimedia Interface (HDMI), and
  - DisplayPort

Connectors of different digital display interfaces. From left to right: VGA, DVI, HDMI, DisplayPort

Digital Visual Interface (DVI)

- Digital Visual Interface (DVI) was developed by the *Digital Display Working Group* (DDWG) for transferring digital video signals, particularly from a computer's video card to a monitor.
- It carries uncompressed digital video and can be configured to support multiple modes, including DVI-D (digital only), DVI-A (analog only), or DVI-I (digital and analog).
- The support for analog connections makes DVI backward compatible with VGA (Video Graphics Array).

- DVI’s digital video transmission format is based on *PanelLink*, a high-speed serial link technology using *transition minimized differential signaling* (TMDS).
- Through DVI, a source, e.g., video card, can read the display’s *extended display identification data* (EDID), which contains the display’s identification, color characteristics, and table of supported video modes.
- When a source and a display are connected, the source first queries the display’s capabilities by reading the monitor’s EDID block.
- A preferred mode or native resolution can then be chosen.
- In a single-link mode, the maximum pixel clock frequency of DVI is 165MHz, which supports a maximum resolution of 2.75megapixels at the 60Hz refresh rate.
- This allows a maximum 16:9 screen resolution of 1,920×1,080 at 60 Hz.

High-Definition Multimedia Interface (HDMI)

- HDMI is a newer digital audio/video interface developed to be backward-compatible with DVI.
- Its electrical specifications, in terms of TMDS and VESA/DDC links, are identical to those of DVI.
- HDMI, however, differs from DVI in the following aspects:
  - HDMI does not carry analog signal and hence is not compatible with VGA.
  - DVI is limited to the RGB color range (0–255). HDMI supports both RGB and YUV 4:4:4 or 4:2:2. The latter are more common in application fields other than computer graphics.
  - HDMI supports digital audio, in addition to digital video.
- The maximum pixel clock rate for HDMI 1.0 is 165MHz, HDMI 1.3 increases that to 340MHz while the latest HDMI 2.0 supports 4K resolution at 60 fps.

DisplayPort

- DisplayPort is the first display interface that uses packetized data transmission, like the Internet or Ethernet.
- Specifically, it is based on small data packets known as *micro packets*, which can embed the clock signal within the data stream.
- DisplayPort can achieve a higher resolution yet with fewer pins than the previous technologies.
- The use of data packets also allows DisplayPort to be extensible.
- DisplayPort can be used to transmit audio and video simultaneously, or either of them.
- It has a much higher video bandwidth, enough for four simultaneous 1080P 60Hz displays, or 4K video at 60 Hz.

- Compared with HDMI, DisplayPort has slightly more bandwidth, which also accommodates multiple streams of audio and video to separate devices.
- It is royalty-free, while HDMI charges an annual fee to manufacturers. These points make DisplayPort a strong competitor to HDMI in the consumer electronics market

Multimedia Systems
Lecture – 15

By

Dr. Priyambada Subudhi
Assistant Professor
IIIT Sri City

3D Video and TV

- Three-dimensional (3D) pictures and movies have been in existence for decades.
- Increasingly, it is in movie theaters, broadcast TV (e.g., sporting events), personal computers, and various handheld devices.
- The main advantage of the 3D video is that it enables the experience of immersion— be there, and really Be there.
- We will see fundamentals of 3D vision or 3D percept, emphasizing stereo vision (or stereopsis) since most modern 3D video and 3D TV are based on stereoscopic vision.

Cues for 3D Percept

- The human vision system is capable of achieving a 3D percept by utilizing multiple cues.
- They are combined to produce optimal (or nearly optimal) depth estimates.
- When the multiple cues agree, this enhances the 3D percept. When they conflict with each other, the 3D percept can be hindered. Sometimes, illusions can arise.

Monocular Cues

- The monocular cues that do not necessarily involve both eyes include:
    - Shading—depth perception by shading and highlights
    - Perspective scaling—converging parallel lines with distance and at infinity
    - Relative size—distant objects appear smaller compared to known same-size objects not in distance
    - Texture gradient—the appearance of textures change when they recede in distance
    - Blur gradient—objects appear sharper at the distance where the eyes are focused, whereas nearer and farther objects are gradually blurred
    - Haze—due to light scattering by the atmosphere, objects at distance have lower contrast and lower color saturation
    - Occlusion—a far object occluded by nearer object(s)
    - Motion parallax—induced by object movement and head movement, such that nearer objects appear to move faster.
- Among the above monocular cues, it has been said that Occlusion and Motion parallax are more effective.

Binocular Cues

- The human vision system utilizes effective binocular vision, i.e., *stereo vision.*
- Our left and right eyes are separated by a small distance, on average approximately 2.5 inches, or 65mm. This is known as the *interocular distance*.
- As a result, the left and right eyes have slightly different views, i.e., images of objects are shifted horizontally.
- The amount of the shift, or *disparity*, is dependent on the object’s distance from the eyes, i.e., its *depth*, thus providing the binocular cue for the 3D percept.
- The horizontal shift is also known as *horizontal parallax*.
- The fusion of the left and right images into single vision occurs in the brain, producing the 3D percept.

Stereo Camera Model

- We can design a simple (artificial) stereo camera system in which the left and right cameras are identical (same lens, same focal length, etc.); the cameras' optical axes are in parallel, pointing at the *Z*-direction, the scene depth.
- The cameras are placed at $(-b/2, 0, 0)$ and $(b/2, 0, 0)$ in the world coordinate system where *b* is camera separation, or the length of the *baseline* where *b* is camera separation, or the length of the *baseline*.
- Given a point $P(X, Y, Z)$ in the 3D space, and x_l and x_r being the *x*-coordinates of its projections on the left and right camera image planes, the following can be derived:

$$d = f\,b/Z,$$

where *f* is the focal length, $d = x_l - x_r$ is the *disparity* or *horizontal parallax*.

- This suggests that disparity *d* is inversely proportional to the depth *Z* of the point *P*.
- Namely, objects near the cameras yield large disparity values, and far objects yield small disparity values. When the point is very far, approaching infinity, $d \to 0$.
- Moreover, objects at the same depth in the scene will have the same disparity *d.* This enables us to depict the 3D space with a stack of *depth planes*, or equivalently, *disparity planes*, which is handy in camera calibration, video processing and analysis.

3DMovie and TV Based on Stereo Vision

- **3D Movie Using Colored Glasses**
- **3D Movies Using Circularly Polarized Glasses**
- **3D TV with Shutter Glasses**

Multimedia Systems
Lecture – 16

By

Dr. Priyambada Subudhi
Assistant Professor
IIIT Sri City

3D Movie Using Colored Glasses

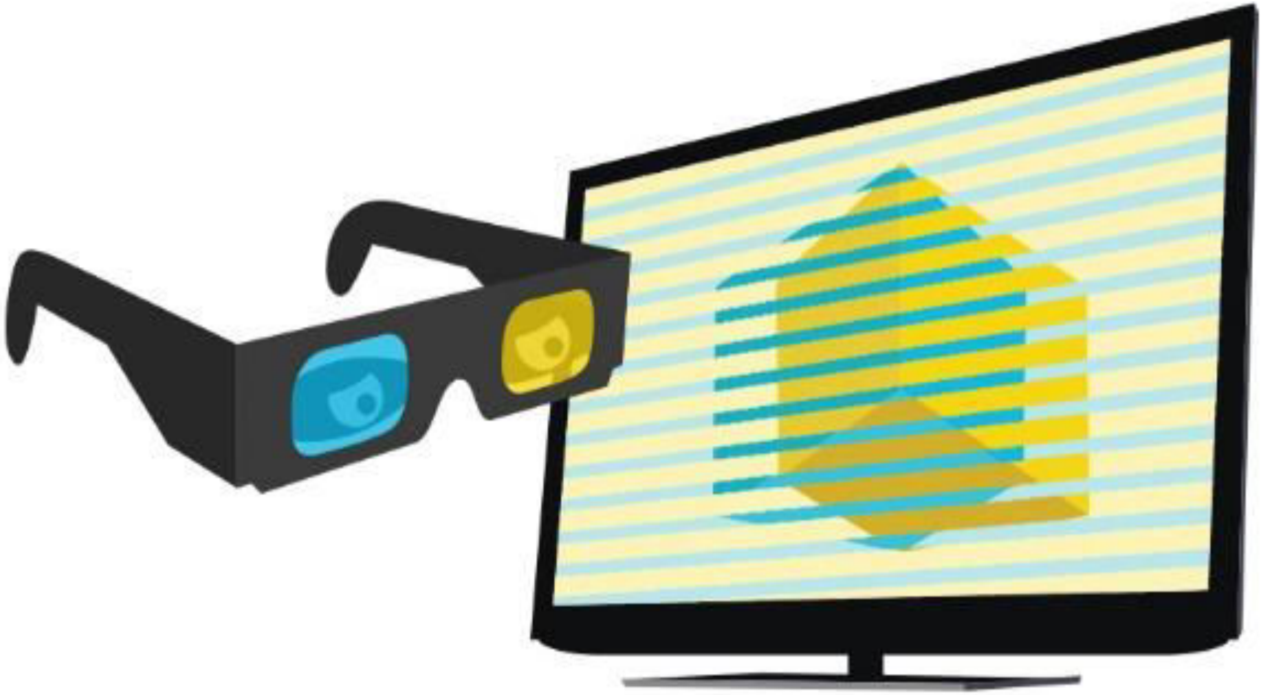
- In the early days, most movie theaters offering a 3D experience provided glasses tinted with complementary colors, usually red on the left and cyan on the right. This technique is called *Anaglyph 3D*.
- Anaglyph 3D images contain two differently filtered colored images, one for each eye. The left image is filtered to remove Blue and Green, and the right image is filtered to remove Red.
- When viewed through the "color-coded" "anaglyph glasses", each of the two images reaches the eye it's intended for, revealing an integrated stereoscopic image.
- The visual cortex of the brain fuses this into the perception of a three-dimensional scene or composition.
- The Anaglyph 3D movies are easy to produce. However, due to the color filtering, the color quality is not necessarily the best.

Anaglyph 3d glasses

3D Movies Using Circularly Polarized Glasses

- Nowadays, the dominant technology in 3D movie theaters is the RealD Cinema System.
- Movie-goers are required to wear polarized glasses in order to see the movie in 3D.
- Basically, the lights from the left and right pictures are polarized in different directions. They are projected and superimposed on the same screen.
- The left and right polarized glasses that the audience wear are polarized accordingly, which allows one of the two polarized pictures to pass through while blocking the other.
- Circularly polarized glasses are used so the users can tilt their heads and look around a bit more freely without losing the 3D percept.

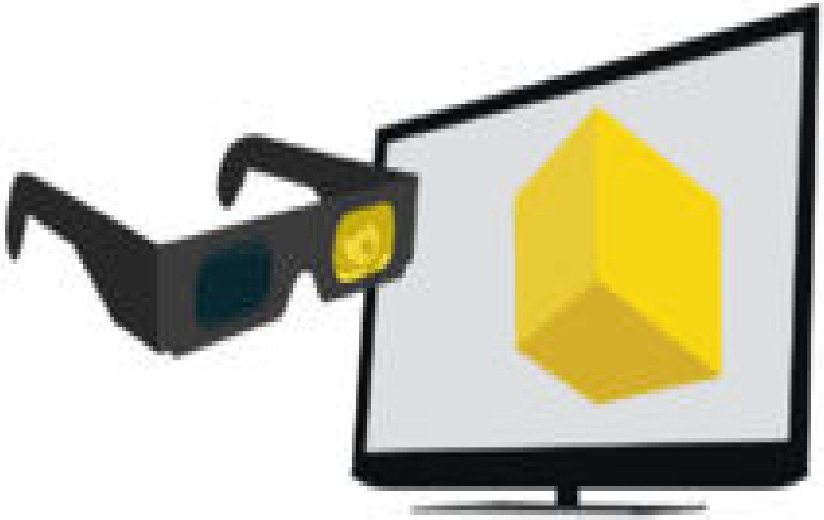
Polarized 3D systems



3D TV with Shutter Glasses

- Most TVs for home entertainment, however, use *Shutter Glasses*.
- Basically, the liquid crystal layer on the glasses that the user wears becomes opaque (behaving like a shutter) when some voltage is applied. It is otherwise transparent.
- The glasses are actively (e.g., via Infra-Red) synchronized with the TV set that alternately shows left and right images (e.g., 120Hz for the left and 120Hz for the Right) in a Time Sequential manner.
- 3D vision with shutter glasses can readily be realized on desktop computers or laptops with a modest addition of specially designed hardware and software. The NVIDIA GeForce 3D Vision Kit is such an example.

A pair of Crystal Eyes shutter glasses



Audio

- Audio information is crucial for multimedia presentations and, in a sense, is the simplest type of multimedia data.

- However, some important differences between audio and image information cannot be ignored. For example, while it is customary and useful to occasionally drop a video frame from a video stream, to facilitate viewing speed, we simply cannot do the same with sound information or all sense will be lost from that dimension.

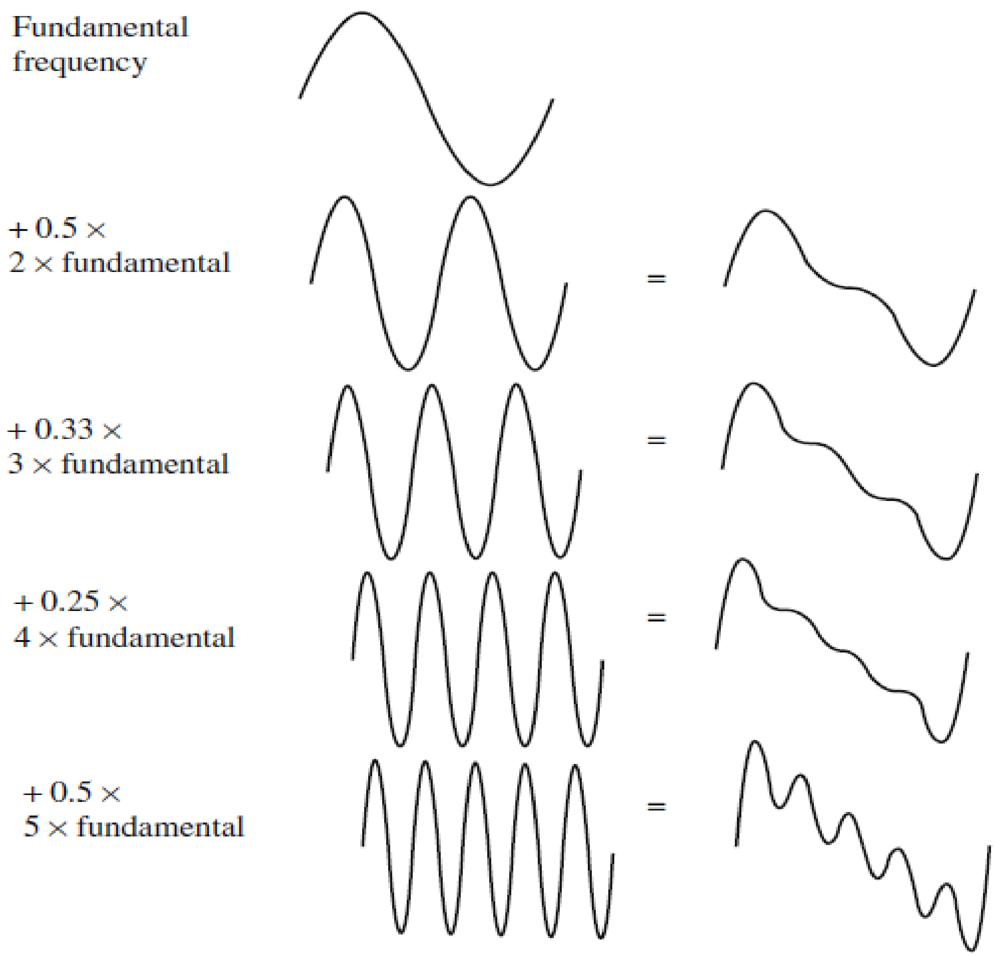
Digitization of Sound

What is Sound?

- Sound is a wave phenomenon like light, but it is macroscopic and involves molecules of air being compressed and expanded under the action of some physical device.
- For example, a speaker in an audio system vibrates back and forth and produces a longitudinal pressure wave that we perceive as sound.
- Without air there is no sound—for example, in space.
- Since sound is a pressure wave, it takes on continuous values, as opposed to digitized ones with a finite range.
- Nevertheless, if we wish to use a digital version of sound waves, we must form digitized representations of audio information.

- Although such pressure waves are longitudinal, they still have ordinary wave properties and behaviors, such as reflection (bouncing), refraction (change of angle when entering a medium with a different density), and diffraction (bending around an obstacle). This makes the design of “surround sound” possible.
- In general, any signal can be decomposed into a sum of sinusoids, if we are willing to use enough sinusoids.
- A weighted sinusoids can build up quite a complex signal.

Building up a complex signal by superposing sinusoids



Multimedia Systems
Lecture – 17

By

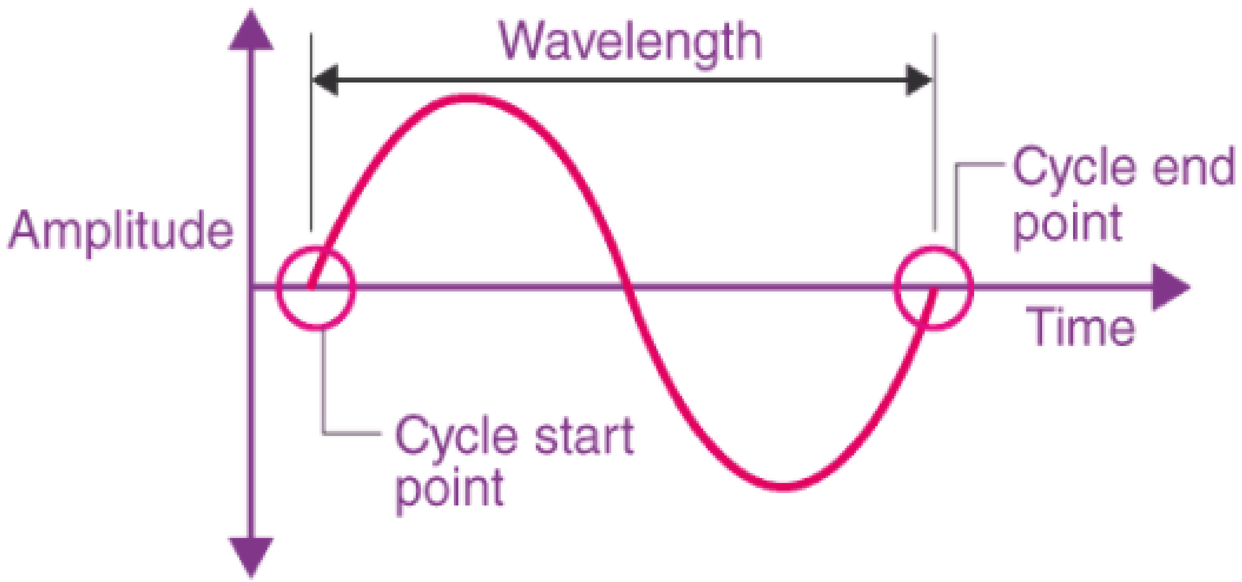
Dr. Priyambada Subudhi
Assistant Professor
IIIT Sri City

Characteristics of Sound

- The sound wave is having the following characteristics
- Amplitude:
  - It refers to the distance of the maximum vertical displacement of the wave from its mean position.
  - In sound, amplitude refers to the magnitude of compression and expansion experienced by the medium the sound wave is travelling through.
  - This amplitude is perceived by our ears as loudness. High amplitude is equivalent to loud sounds.
- Wavelength:
  - A sound wave is made of areas of high pressure alternated by an area of low pressure.
  - The high-pressure areas are represented as the peaks of the graph. The low-pressure areas are represented as troughs of the graph.
  - The physical distance between two consecutive peaks in a sound wave is referred to as the wavelength of the sound wave.

- Frequency/ Pitch of the Sound Waves
  - Frequency in a sound wave refers to the rate of the vibration of the sound travelling through the air. This parameter decides whether a sound is perceived as high pitched or low pitched.
  - In sound, the frequency is also known as **Pitch**.
  - The frequency of the vibrating source of sound is calculated in cycles per second (Hertz).

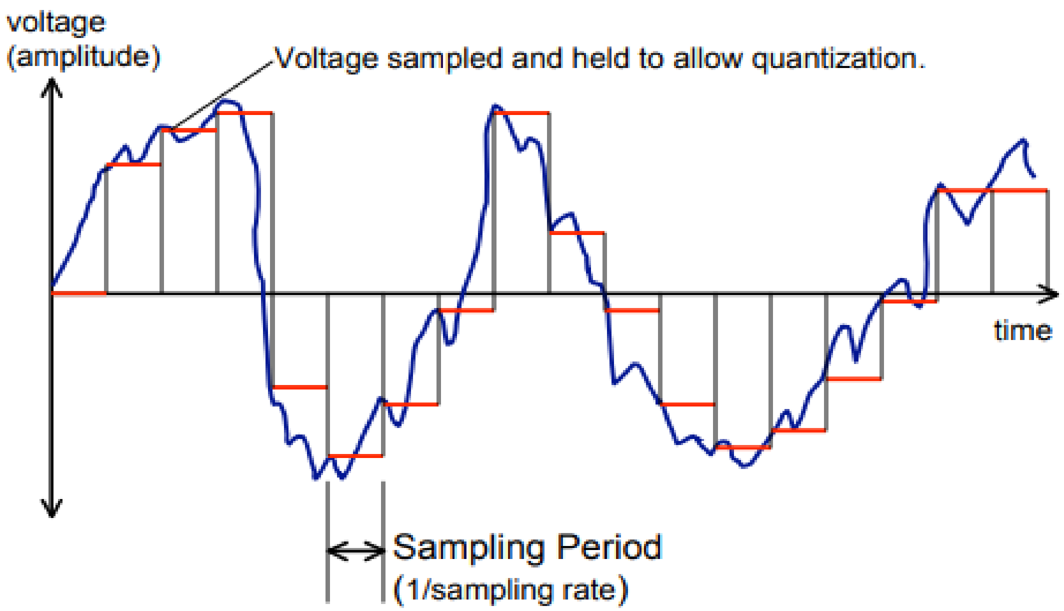
A depiction of Sound Waves in Waveform

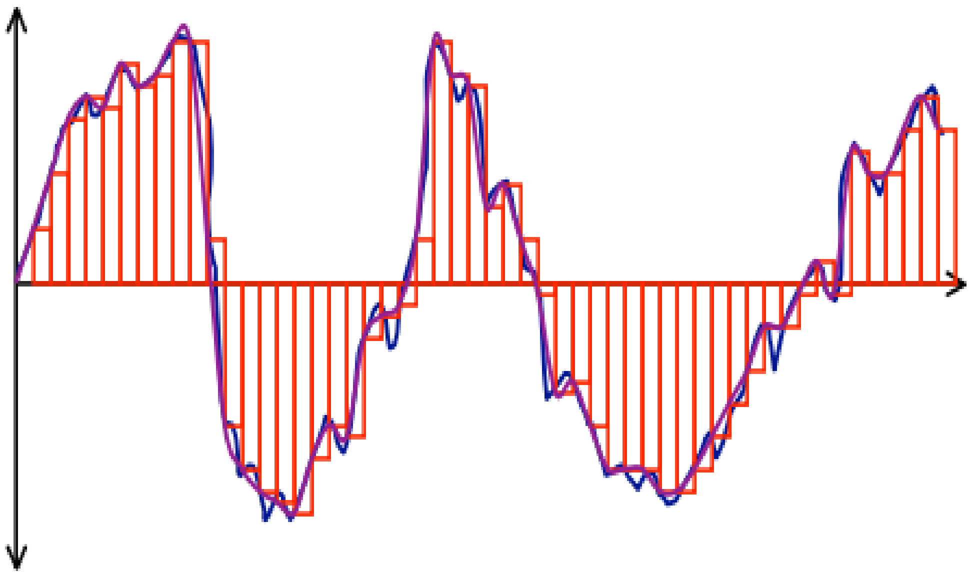


Digitization

- Since there is only one independent variable in sound i.e. time, we call this a 1D signal—as opposed to images, with data that depends on two variables, *x*, and *y.*
- The amplitude value is a continuous quantity. To fully digitize the sound signal, we have to *sample* in time and in amplitude.
- Sampling means measuring the quantity we are interested in, usually at evenly spaced intervals.
- The first kind of sampling—using measurements only at evenly spaced *time* intervals—is simply called *sampling* and the rate at which it is performed is called the *sampling rate* or *sampling frequency*.

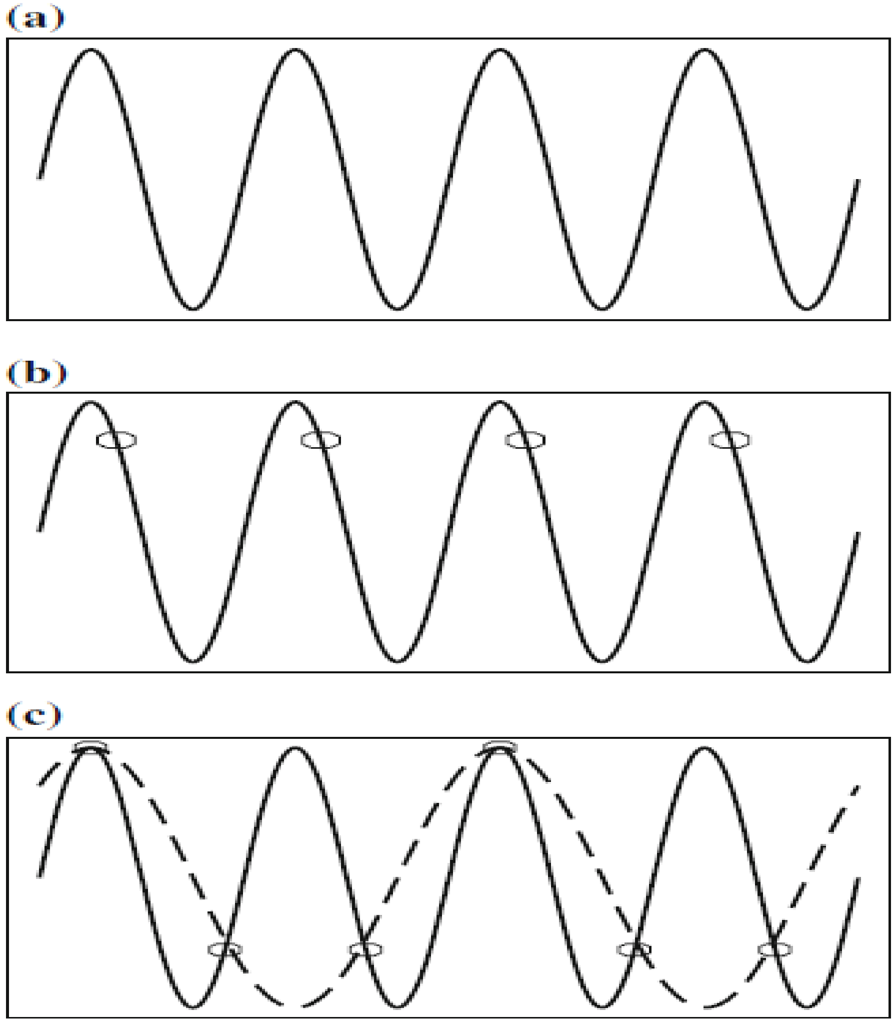
Sampling


voltage
(amplitude)
Voltage sampled and held to allow quantization.
time
Sampling Period
(1/sampling rate)



- For audio, typical sampling rates are from 8 kHz (8,000 samples per second) to 48 kHz.
- The human ear can hear from about *20 Hz to as much as 20 kHz*; above this level, we enter the range of ultrasound.
- The human voice can reach approximately 4 kHz.
- *Nyquist sampling rate :*
  - To preserve the full information in the signal, it is necessary to sample at twice the maximum frequency of the signal. This is known as the Nyquist rate.
  - If we sample the signal at a frequency that is lower that the Nyquist rate, when the signal is converted back into a continuous time signal, it will exhibit a phenomenon called ***aliasing***. Aliasing is the presence of unwanted components in the reconstructed signal.

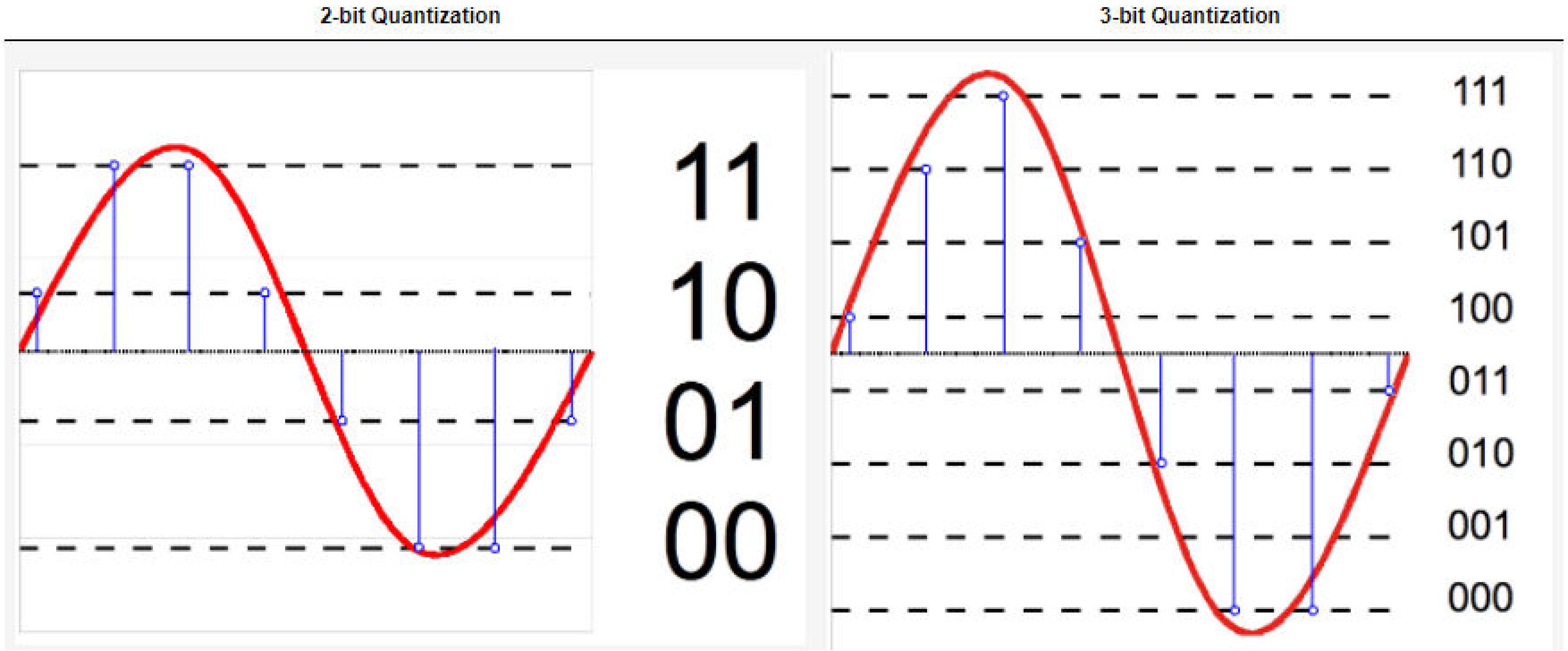
Aliasing: **a)** a single frequency; **b)** sampling at exactly the frequency produces a constant; **c)** sampling at 1.5 times per cycle produces an *alias* frequency that is perceived



Quantization:

- Sampling in the amplitude or voltage dimension is called *quantization or It* refers to the process of transforming a sampled analog signal, to a digital signal, which has a discrete set of values.
- While we have discussed only uniform sampling, with equally spaced sampling intervals, non-uniform sampling is possible. This is not used for sampling in time but is used for quantization.
- Typical uniform quantization rates are 8-bit and 16-bit; 8-bit quantization divides the vertical axis into 256 levels, and 16-bit divides it into 65,536 levels.
- Quantization Error: A digitized sample can have a maximum error of one-half the discretization step size.

2-bit and 3-bit Quantization



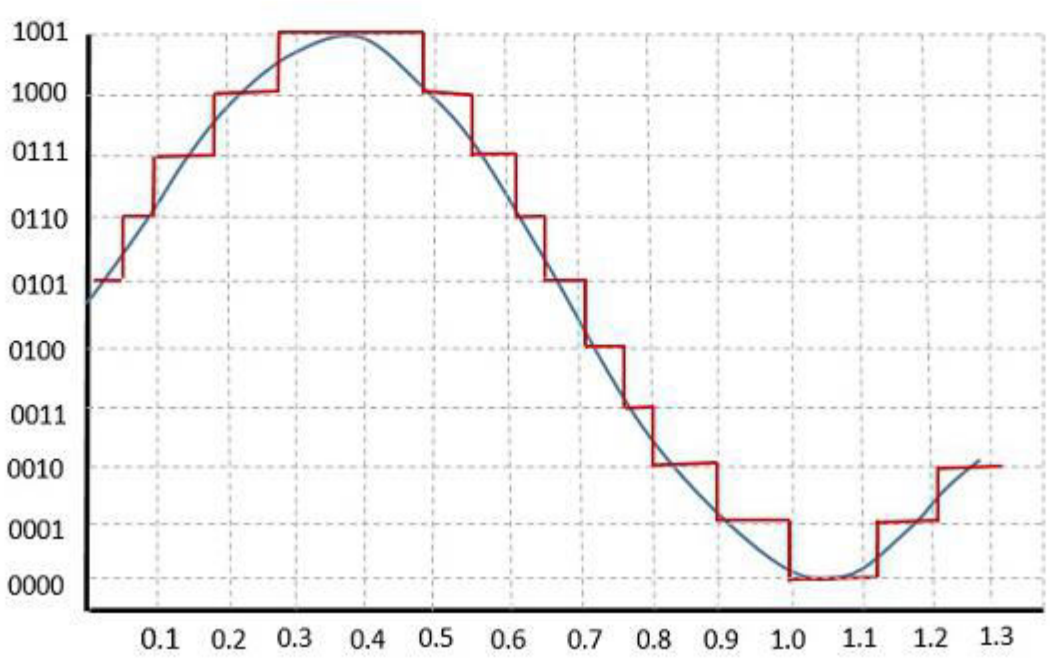
Multimedia Systems
Lecture – 18

By

Dr. Priyambada Subudhi
Assistant Professor
IIIT Sri City

Linear and Nonlinear/ Uniform and Non-uniform Quantization

- Samples are typically stored as uniformly quantized values. This is called *linear or uniform quantization.*



- There are two types of uniform quantization. They are *Mid-Rise* type and *Mid-Tread* type.
- The **Mid-Rise** type is so called because the origin lies in the middle of a raising part of the stair-case like graph. The quantization levels in this type are even in number.
- The **Mid-tread** type is so called because the origin lies in the middle of a tread of the stair-case like graph. The quantization levels in this type are odd in number.
- The difference between an input value and its quantized value is called a **Quantization Error**.

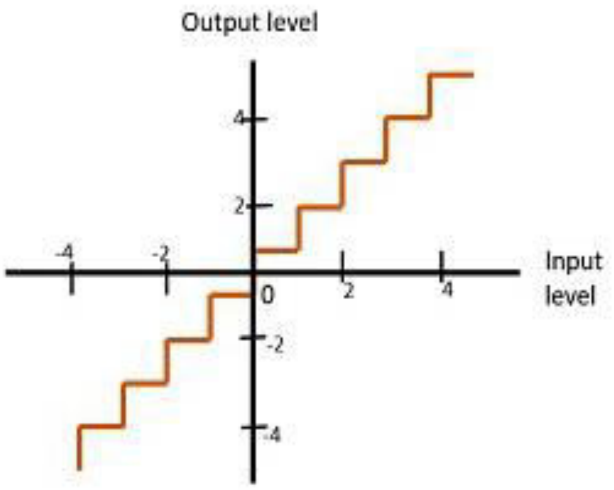


Fig 1 : Mid-Rise type Uniform Quantization

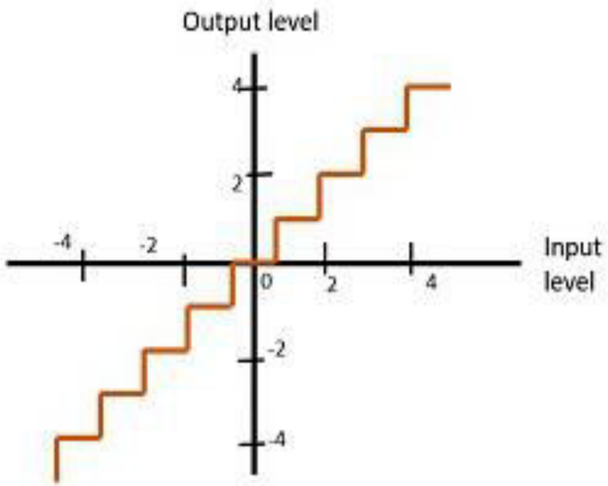


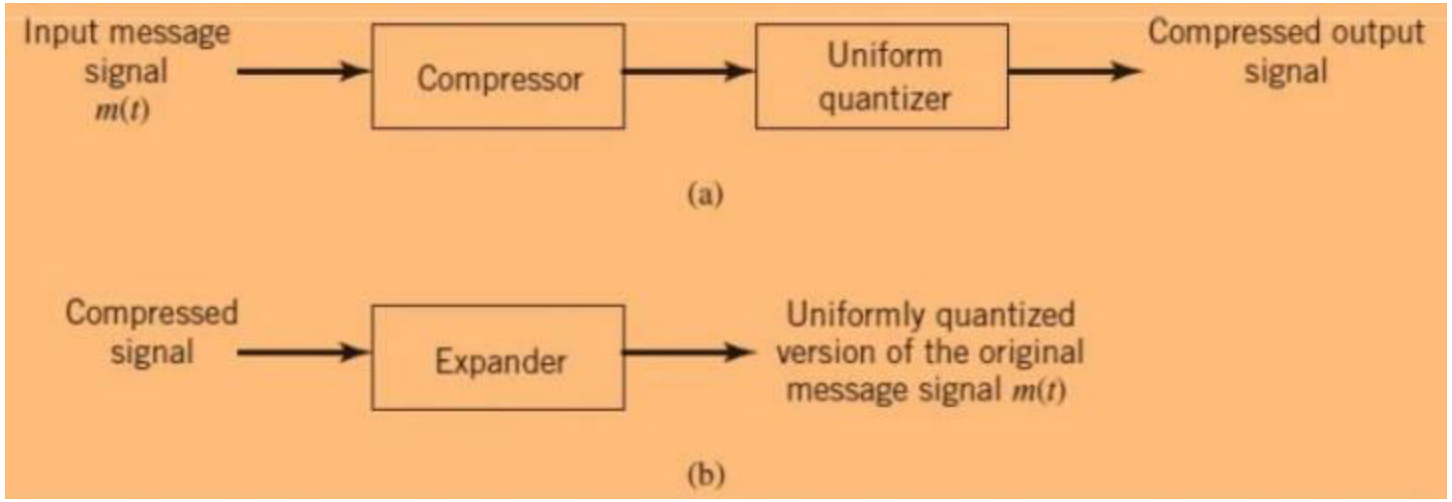
Fig 2 : Mid-Tread type Uniform Quantization

Non-uniform Quantization

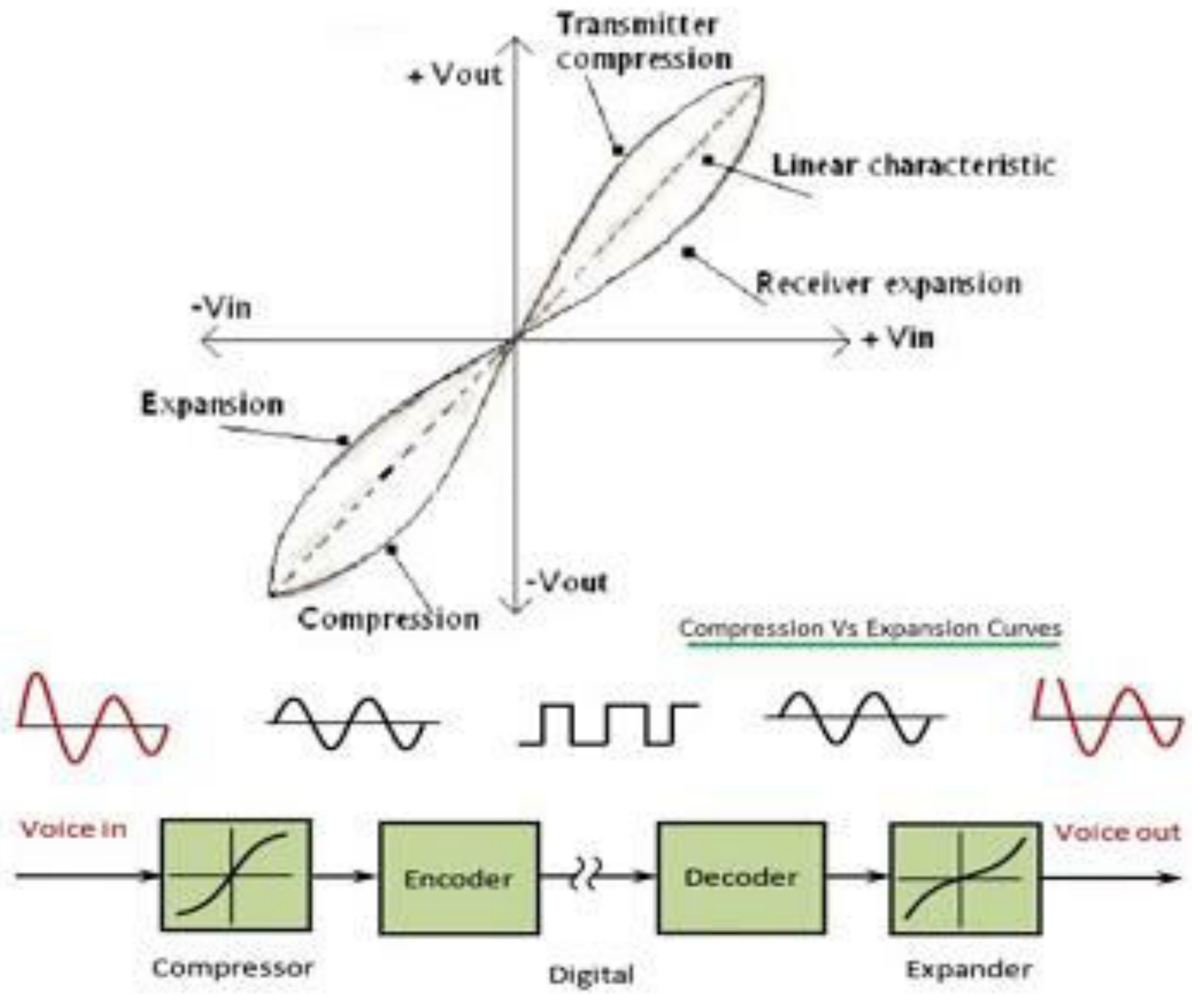
- If the quantization characteristic is nonlinear then the step size is not constant and quantization is known as non-uniform quantization.
- It is mostly used in case of speech or music as here the variation in amplitude is high which is expressed as crest factor and is given by

  crest factor = peak value of signal/rms value of signal
- Non-uniform quantization is achieved using companding.

- <u>Companding:</u>
- It is derived from two words, ***Compressing*** and ***Expanding***.
- The desired form of non-uniform quantization can be achieved by using compressor followed by a uniform quantizer.



Companding Process



μ-law and A-law companding

- μ-law is popular technique used in USA and Japan.
- Here the input and output relationship is given by

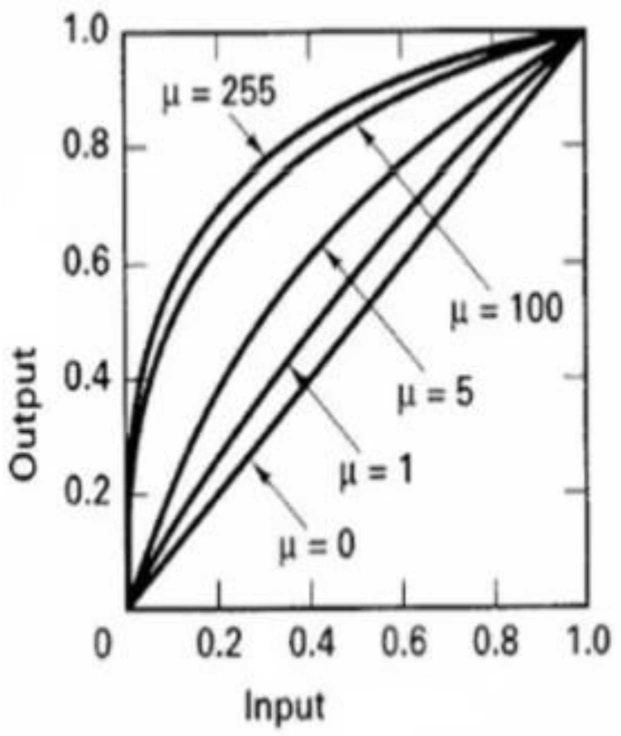
$$r = \frac{\text{sign}(s)}{\ln(1+\mu)} \ln\left\{1 + \mu \left|\frac{s}{s_p}\right|\right\}, \qquad \left|\frac{s}{s_p}\right| \le 1$$

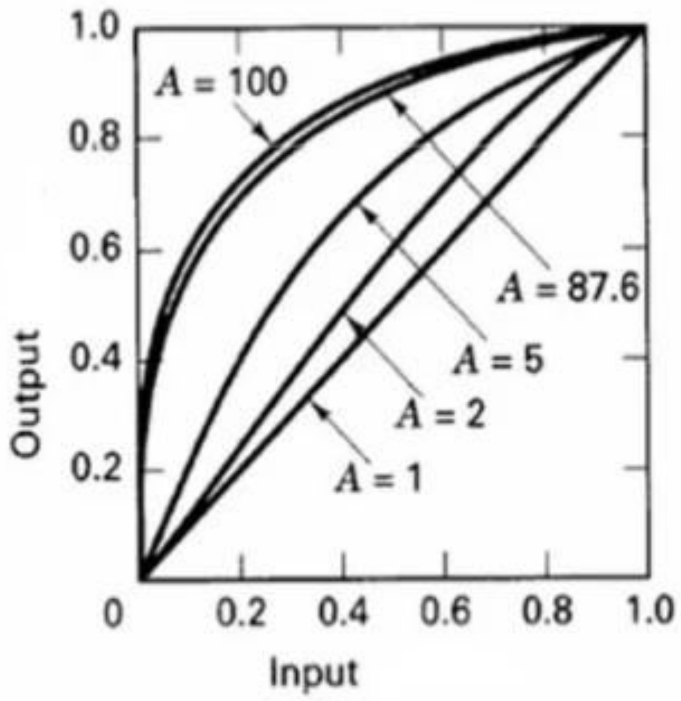
- A very similar rule, called *A-law*, is used in telephony in Europe.

$$r = \begin{cases} \frac{A}{1+\ln A}\left(\frac{s}{s_p}\right), & \left|\frac{s}{s_p}\right| \le \frac{1}{A} \\ \frac{\text{sign}(s)}{1+\ln A}\left[1 + \ln A\left|\frac{s}{s_p}\right|\right], & \frac{1}{A} \le \left|\frac{s}{s_p}\right| \le 1 \end{cases}$$

$$\text{where } \ \text{sign}(s) = \begin{cases} 1 & \text{if } s > 0, \\ -1 & \text{otherwise} \end{cases}$$

μ-law and A-law Compression Characteristics

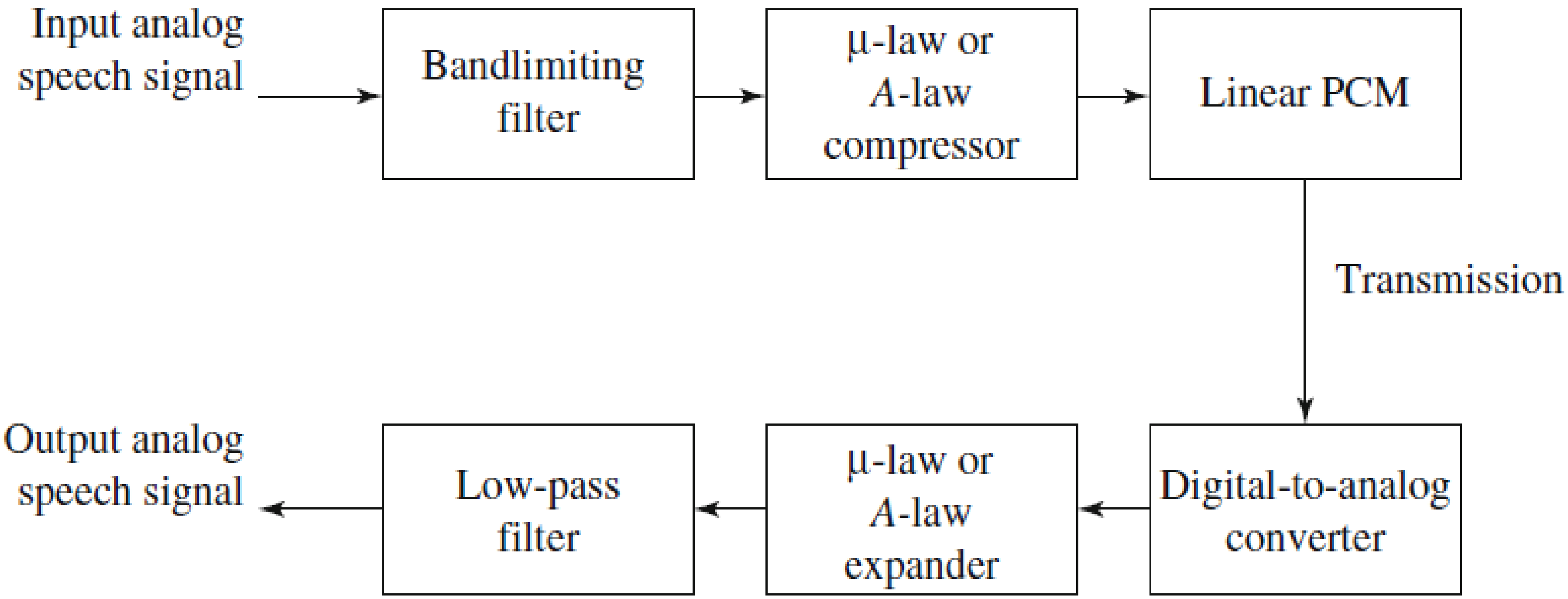




Pulse Code Modulation (PCM)

- **Modulation** is the process of varying one or more parameters of a carrier signal in accordance with the instantaneous values of the message signal.
- There are many modulation techniques, which are classified according to the type of modulation employed. Of them all, the digital modulation technique used is **Pulse Code Modulation.**
- We know that the basic techniques for creating digital signals from analog ones consist of *sampling* and *quantization*.
- Pulse Code Modulation, is a formal term for the sampling and quantization we have already been using.
- *Pulse* comes from an engineer's point of view that the resulting digital signals can be thought of as infinitely narrow vertical "pulses."

Basic Elements of PCM



Multimedia Systems Lecture – 19

By

Dr. Priyambada Subudhi
Assistant Professor
IIIT Sri City

Differential Pulse Code Modulation (DPCM)

- It is based on differential predictive coding.
- Audio is often stored not in simple PCM but in a form that exploits differences.
- Generally, if a time-dependent signal has some consistency over time (*temporal redundancy*), the difference signal—subtracting the current sample from the previous one—will have a more peaked histogram, with a maximum around zero.
- For a start, differences will generally be smaller numbers and hence offer the possibility of using fewer bits to store.

- Suppose our integer sample values are in the range 0 .. 255. Then differences could be as much as −255 .. 255. So we have unfortunately increased our *dynamic range* (ratio of maximum to minimum) by a factor of two.
- Let's formalize our statement of what we are doing by defining the integer signal as the set of values f_n.
- Then we *predict* $\hat{f}_n$ values as simply the previous value, and we define the error e_n as the difference between the actual and predicted signals:

$$\hat{f}_n = f_{n-1}$$
$$e_n = f_n - \hat{f}_n$$

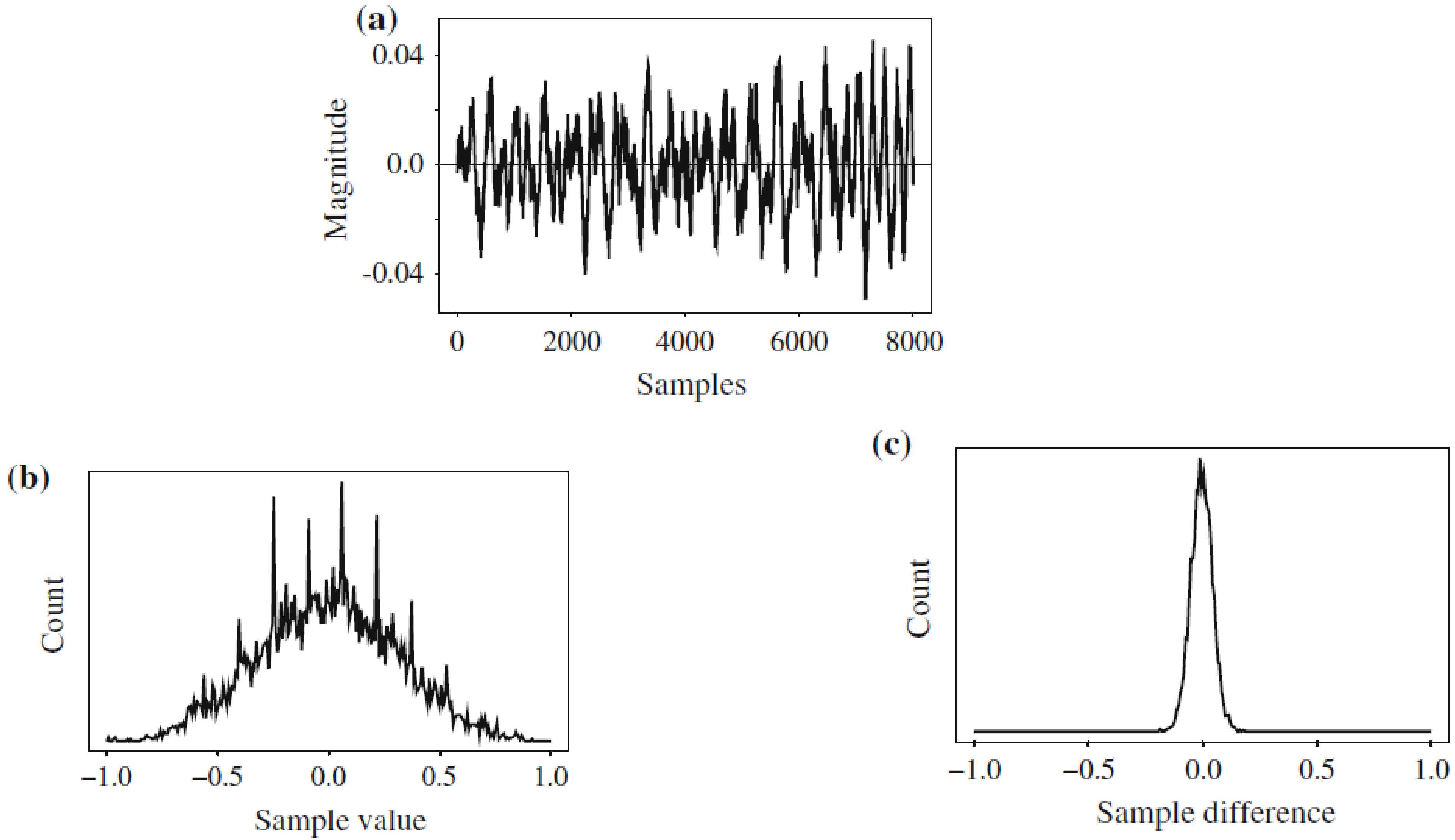
- We certainly would like our error value e_n to be as small as possible.
- Therefore, we would wish our prediction $\hat{f}_n$ to be as close as possible to the actual signal f_n.

- But for a particular sequence of signal values, some *function* of a few of the previous values, f_{n-1}, f_{n-2}, f_{n-3}, etc., may provide a better prediction of f_n.

$$\hat{f}_{\mathrm{n}} = \sum_{k=1}^{2 \text{ to } 4} a_{n-k} f_{n-k}$$

- The idea of forming differences is to make the histogram of sample values more peaked.
- A plot of 1 second of sampled speech at 8 kHz, with magnitude resolution of 8 bits per sample is considered.
- A histogram of these values is centered around zero. However, the histogram for corresponding speech signal *differences*: difference values are much more clustered around zero than are sample values themselves.
- So we can assign short codes to prevalent values like zeros here and long codewords to rarely occurring ones.

Differencing concentrates the histogram: **a** digital speech signal; **b** histogram of digital speech signal values; **c** histogram of digital speech signal differences



- Suppose samples are in the range 0 .. 255, and differences are in −255 .. 255. Then
- Define SU and SD as shifts by 32.
- Then we could in fact produce code-words for a limited set of signal differences, say only the range −15 .. 16.
- Differences (that inherently are in the range −255 .. 255) lying in the limited range can be coded as is, but if we add the extra two values for SU, SD, a value outside the range −15 .. 16 can be transmitted as a series of shifts, followed by a value that is indeed inside the range −15 .. 16.
- For example, 100 is transmitted as SU, SU, SU, 4, where (the codes for) SU and for 4 are what are sent.

- Example: suppose we devise a predictor for $\hat{f}_n$ as follows:

$$\hat{f}_n = \lfloor \frac{1}{2}(f_{n-1} + f_{n-2}) \rfloor$$
$$e_n = f_n - \hat{f}_n$$

- Then the error e_n (or a codeword for it) is what is actually transmitted. Suppose we wish to code the sequence $f1, f2, f3, f4, f5$ = 21, 22, 27, 25, 22.
- For the purposes of the predictor, we'll invent an extra signal value f_0, equal to $f1$ = 21, and first transmit this initial value, uncoded; after all, every coding scheme has the extra expense of some header information.

- Then the first error, *e*1, is zero, and subsequently

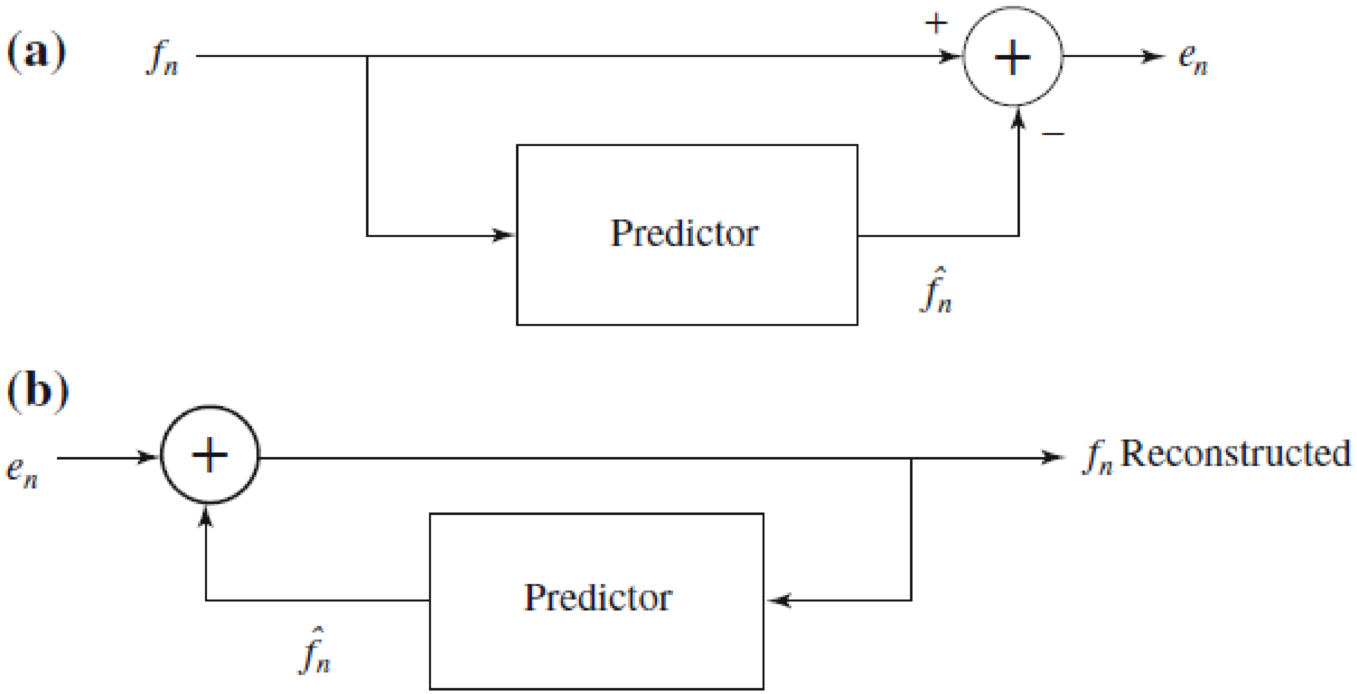
$$\hat{f}_2 = 21, \quad e_2 = 22 - 21 = 1$$

$$\hat{f}_3 = \lfloor \frac{1}{2}(f_2 + f_1) \rfloor = \lfloor \frac{1}{2}(22 + 21) \rfloor = 21$$
$$e_3 = 27 - 21 = 6$$

$$\hat{f}_4 = \lfloor \frac{1}{2}(f_3 + f_2) \rfloor = \lfloor \frac{1}{2}(27 + 22) \rfloor = 24$$
$$e_4 = 25 - 24 = 1$$

$$\hat{f}_5 = \lfloor \frac{1}{2}(f_4 + f_3) \rfloor = \lfloor \frac{1}{2}(25 + 27) \rfloor = 26$$
$$e_5 = 22 - 26 = -4$$

Schematic diagram for Predictive Coding: **a** encoder; **b** decoder



Multimedia Systems
Lecture – 20

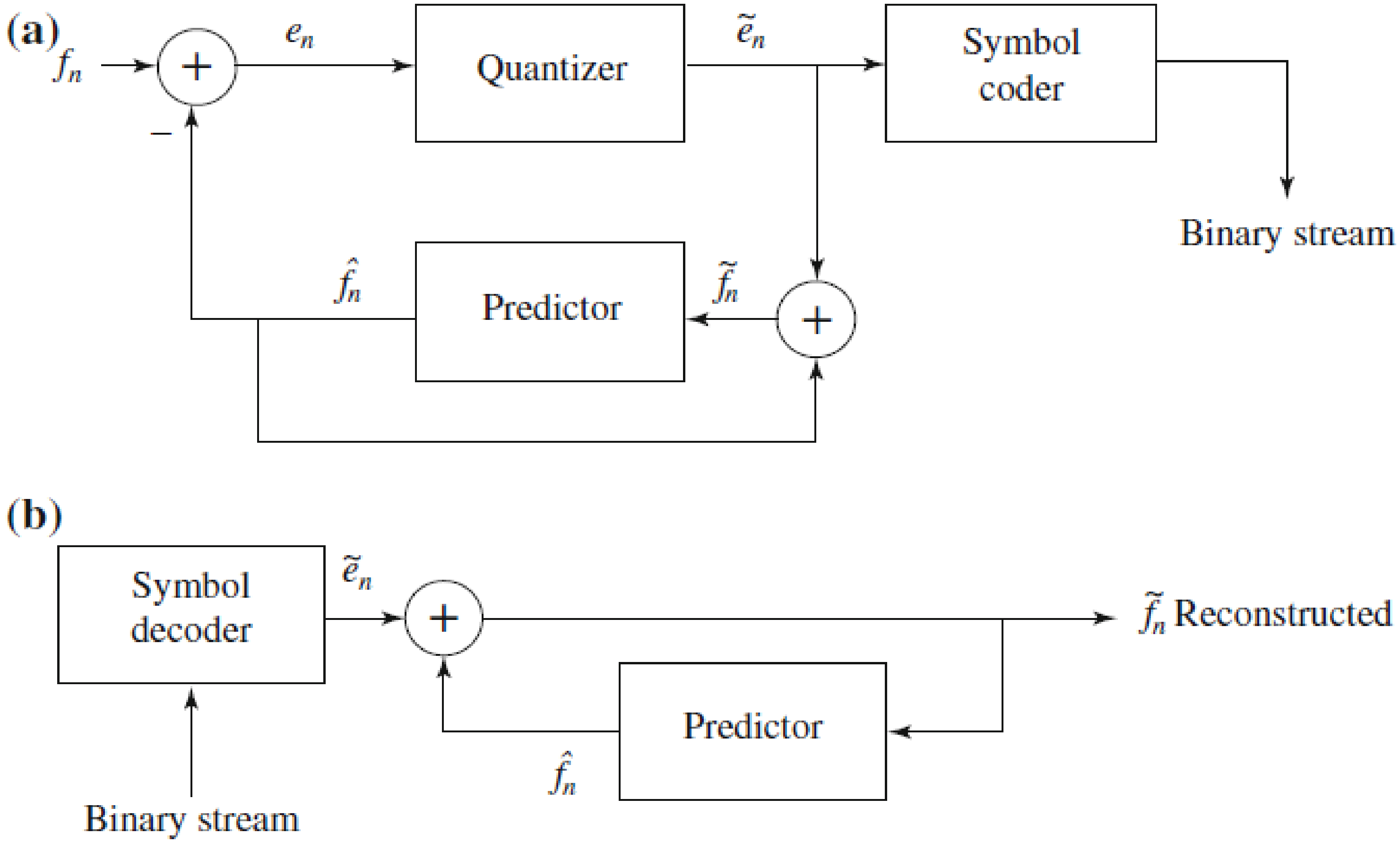
By

Dr. Priyambada Subudhi
Assistant Professor
IIIT Sri City

- Differential Pulse Code Modulation (DPCM) is exactly the same as Predictive Coding, Predictive coding except that it incorporates a quantizer step.
- We shall call the original signal f_n, the predicted signal $\hat{f}_n$, and the quantized, reconstructed signal $\tilde{f}_n$. How DPCM operates is to form the prediction, from an error e_n by subtracting the prediction from the actual signal, then quantize the error to a quantized version, $\tilde{e}_n$.
- The equations that describe DPCM are as follows

$$
\begin{aligned}
\hat{f}_n &= \text{function\_of}\ (\tilde{f}_{n-1}, \tilde{f}_{n-2}, \tilde{f}_{n-3}, \ldots) \\
e_n &= f_n - \hat{f}_n \\
\tilde{e}_n &= Q[e_n] \\
&\quad \text{transmit codeword}(\tilde{e}_n) \\
&\quad \text{reconstruct: } \tilde{f}_n = \hat{f}_n + \tilde{e}_n
\end{aligned}
$$

Schematic diagram for DPCM: **a** encoder; **b** decoder



- Codewords for quantized error values $\tilde{e}_n$ are produced using entropy coding, such as Huffman coding.
- Notice that the predictor is always based on the reconstructed, quantized version of the signal $\tilde{f}_n$: the reason for this is that then the encoder side is not using any information not available to the decoder side.
- The main effect of the coder–decoder process is to produce $\tilde{f}_n$ reconstructed, quantized signal values $\tilde{f}_n = \hat{f}_n + \tilde{e}_n$
- The distortion is the average squared error $[\sum_{n=1}^{N}(\tilde{f}_n - f_n)^2]/N$
- The predictor makes use of the reconstructed, quantized signal values not actual signal values f_n—that is, the encoder simulates the decoder in the predictor path. The quantizer can be uniform or nonuniform.

- The prediction value $\hat{f}_n$ is based on however much history the prediction scheme requires: we need to buffer previous values of $\tilde{f}_n$ to form the prediction.
- Notice that the quantization noise, $f_n - \tilde{f}_n$ is equal to the quantization effect on the error term, $e_n - \tilde{e}_n$.
- Example: Suppose we adopt a particular predictor as

$$\hat{f}_n = \text{trunc}\left[\left(\tilde{f}_{n-1} + \tilde{f}_{n-2}\right)/2\right]$$

so that $e_n = f_n - \hat{f}_n$ is an integer.

- the particular quantization scheme

$$\tilde{e}_n = Q[e_n] = 16 * \text{trunc}\left[(255 + e_n)/16\right] - 256 + 8$$
$$\tilde{f}_n = \hat{f}_n + \tilde{e}_n$$

- Suppose we wish to code the sequence $f1, f2, f3, f4, f5$ = 130, 150, 140, 200, 230.

- We prepend extra values f_0 = 130 in the datastream that replicate the first value, f_1, and initialize with quantized error $\tilde{e}_1 \equiv 0$, so that we ensure the first reconstructed value is exact: $\tilde{f}_1 = 130$.
- Then subsequent values calculated are as follows

$$
\begin{aligned}
\hat{f} &= \boxed{130},\ 130,\ 142,\ 144,\ 167 \\
e &= \boxed{0},\ 20,\ -2,\ 56,\ 63 \\
\tilde{e} &= \boxed{0},\ 24,\ -8,\ 56,\ 56 \\
\tilde{f} &= \boxed{130},\ 154,\ 134,\ 200,\ 223
\end{aligned}
$$

Delta Modulation (DM)

- It is a much-simplified version of DPCM often used as a quick analog-to-digital converter.

Uniform-Delta DM

- The idea in DM is to use only a *single* quantized error value, either positive or negative. Such a 1-bit coder thus produces coded output that follows the original signal in a staircase fashion.
- The relevant set of equations is as follows:

$$\hat{f}_n = \tilde{f}_{n-1}$$

$$e_n = f_n - \hat{f}_n = f_n - \tilde{f}_{n-1}$$

$$\tilde{e}_n = \begin{cases} +k \text{ if } e_n > 0, & \text{where } k \text{ is a constant} \\ -k \text{ otherwise,} & \end{cases}$$

$$\tilde{f}_n = \hat{f}_n + \tilde{e}_n$$

- Note that the prediction simply involves a delay.
- Example: Suppose signal values are as follows

f_1	f_2	f_3	f_4
10	11	13	15

- We also define an exact reconstructed value $\tilde{f}_1 = f_1 = 10$.
- Suppose we use a step value *k* = 4. Then we arrive at the following values:

$$\begin{array}{llll} \hat{f}_2 = 10, & e_2 = 11 - 10 = 1, & \tilde{e}_2 = 4, & \tilde{f}_2 = 10 + 4 = 14 \\ \hat{f}_3 = 14, & e_3 = 13 - 14 = -1, & \tilde{e}_3 = -4, & \tilde{f}_3 = 14 - 4 = 10 \\ \hat{f}_4 = 10, & e_4 = 15 - 10 = 5, & \tilde{e}_4 = 4, & \tilde{f}_4 = 10 + 4 = 14 \end{array}$$

- We see that the reconstructed set of values 10, 14, 10, 14 never strays far from the correct set 10, 11, 13, 15.
- It is not difficult to discover that DM copes well with more or less constant signals, but not as well with rapidly changing signals.

Adaptive DM

- However, if the slope of the actual signal curve is high, the staircase approximation cannot keep up.
- A straightforward approach to dealing with a steep curve is to simply change the step size *k adaptively*—that is, in response to the signal's current properties.

Commonly Used Audio Formats

- Digital audio formats emerged with the use and distribution of CD audio discs. These were uncompressed pulse code modulated digital signals in mono and in stereo (*Mono signals are recorded and played back using a single audio channel, while stereo sounds are recorded and played back using two audio channels*).
- However, a number of formats have now become mainstream with the need for streaming, mobile, and surround sound technologies (**Surround sound** is a technique for enriching the fidelity and depth of sound reproduction by using multiple audio channels from speakers that surround the listener (surround channels). Its first application was in movie theaters).

File suffix or logo	Filename	File type	Features
.wav	WAV	Uncompressed PCM coded	Default standard for audio on PCs. WAV files are coded in PCM format.
.au	G.711 μ-law, or ITU μ-law	Uncompressed audio	Universal support for telephone. Packs each 16-bit sample into 8 bits, by using logarithmic table to encode with a 13-bit dynamic range. Encoding and decoding is very fast.
GSM 06.10	Global System for Mobile Communication	Lossy Compressed mobile audio	International standard for cellular telephone technology. Uses linear predictive coding to substantially compress the data. Compression/decompression is slow. Freely available and, thus, widely used
.mp3	MPEG1 Layer3	Compressed audio file format	Uses psychoacoustics for compression Very good bandwidth savings and, hence, used for streaming and Internet downloads.

.ra	Real Audio	Compressed format	Proprietary to Real Audio. Capable of streaming and downloading. Comparable quality to mp3 at high data rates but not so at low data rates
AAC	Advanced Audio Codec MPEG4	Compressed format	Superior quality to .mp3.
.mid	MIDI—Musical Instrument Digital Interface	Descriptive format	MIDI is a language of communication among musical instruments. Description achieved by frequencies, decays, transients, and event lists. Sound has to be synthesized by the instrument.

Multimedia Systems
Lecture – 21

By

Dr. Priyambada Subudhi

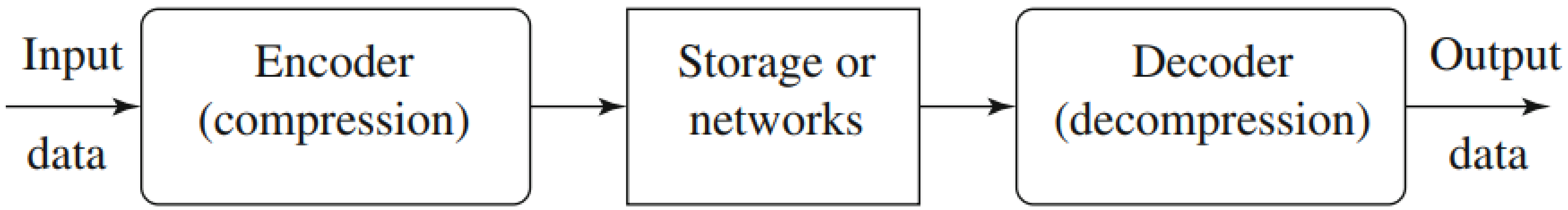
Assistant Professor

IIIT Sri City

Multimedia Data Compression

- The amount of digital media data that is produced in the form of text, video, audio, 3D graphics, and combinations of these media types is extraordinarily large, and the rate of creation increases every day.
- This growing mass of data needs to be stored, accessed, and delivered to a multitude of clients over digital networks, which have varying bandwidths.
- The existence of voluminous data, from creation to storage and delivery, motivates the need for compression.
- The role played in multimedia by data compression, perhaps the most important enabling technology that makes modern multimedia systems possible.

- In a general data compression scheme, in which compression is performed by an encoder and decompression is performed by a decoder. general data compression scheme.



- We call the output of the encoder *codes* or *codewords*.
- The intermediate medium could either be data storage or a communication /computer network.
- If the compression and decompression processes induce no information loss, the compression scheme is *lossless*; otherwise, it is *lossy*.

- Compression Ratio: If the total number of bits required to represent the data before compression is B0 and the total number of bits required to represent the data after compression is B1, then we define the compression ratio as

$$\text{compression ratio} = \frac{B_0}{B_1}.$$

- In general, we would desire any codec (encoder/decoder scheme) to have a compression ratio much larger than 1.0.
- The higher the compression ratio, the better the lossless compression scheme, as long as it is computationally feasible.

Basics of Information Theory

- When transmitting information from a source to a destination, information theory concerns itself with the efficiency and reliability of the transmission.
- Information theory allows us to describe the process to compress data without losing its information content.
- The data could represent virtually anything— simple text, documents, images, graphics, and even binary executables.
- Here we will consider all these varied data types as generic data represented digitally in a binary form by a sequence of 1s and 0s.

- To understand compression, it is necessary to understand the information content of a message.
- Information relates to the organization in the data as a sequence of symbols. If the sequence changes, so does the information.
- For example, you might have a binary data stream that consists of 80,000 bits. These bits might not need to be treated individually, but might instead be grouped into symbols. For instance, if it is known that bits represent a gray-intensity image with each pixel represented by 8 bits, then we have 10,000 pixels or symbols. Furthermore, if width and height of the image are both 100, the bit stream might be interpreted as a gray image of size 100×100. Here, each symbol is represented by 8 bits, and there can be $2^8 = 256$ different possible gray levels or symbols.
- Also, the arrangement of the symbols is important

- Alphabet and Symbol: Information can be thought of as an organization of individual elements called *symbols*.
- An *alphabet* is defined as a distinct and nonempty set of symbols.
- The number or length of symbols in the set is known as the *vocabulary*.
- We can define an alphabet of symbols as S = {s1, s2, s3, s4 . . . sn}. Though n can be very large, in practice, an alphabet is limited and finite and, hence, has a well-defined vocabulary.
- In the previous example involving a gray image, each pixel is represented by 8 bits and can have one of 2^8 or 256 unique values. Here, vocabulary consists of 256 symbols, where each symbol is represented or coded by 8 bits.
- Sequence: A series of symbols of a given alphabet form a sequence. For the alphabet {s1, s2, s3, s4}, a sample sequence is s1 s2 s1 s2 s2 s2 s1 s2 s3 s4 s1 s2 s3.
- A sequence of symbols is also termed a message produced by the source using the alphabet, and represents information.
- When looking at a sequence, some symbols might occur more commonly than other symbols. The frequency of occurrence of a symbol is an important factor when coding information represented by the symbols. The frequency is also known as probability.

- Symbol Probability: The probability of occurrence of a symbol is defined by the ratio of the number of occurrences of that symbol over the length of the entire message.

$$P_i = \frac{m_i}{N},$$

- where m_i is the number of times symbol s_i occurs in the message of length N.
- Most coding algorithms make extensive use of symbol probabilities to obtain optimal codes for compression.

- Entropy: Shannon's information theory borrows the definition of entropy from physics to quantify the amount of information contained in a message of symbols given their probabilities of occurrence.
- For source-producing symbols, where each symbol i has a probability distribution P_i , the entropy is defined as

$$H = \sum P_i \log_2\left(\frac{1}{P_i}\right) = -\sum P_i \log_2 P_i.$$

- For the symbol s_i having a probability P_i , Shannon defined the notion of self-information of the symbol given by $\log_2(1/P_i)$.
- The self-information represents number of bits of information contained in the symbol and, hence, the number of bits used to send that message.
- Entropy, then, becomes the weighted average of the information carried by each symbol and, hence, the average symbol length.

- For example, if the probability of having the character n in a manuscript is 1/32, the amount of information associated with receiving this character is 5 bits.
- In other words, a character string nnn will require 15 bits to code.
- The definition of entropy is aimed at identifying often-occurring symbols in the datastream as good candidates for short codewords in the compressed bitstream.
- We use a variable-length coding scheme for entropy coding—frequently occurring symbols are given codes that are quickly transmitted, while infrequently occurring ones are given longer codes.

- if the information source S is a gray-level digital image, each si is a gray-level intensity ranging from 0 to (2k − 1), where k is the number of bits used to represent each pixel in an uncompressed image. The range is often [0, 255], since 8 bits are typically used.
- Fig. a shows the histogram of an image with uniform distribution of gray-level intensities—that is, ∀i p_i = 1/256. Hence, the entropy of this image is

$$\eta = \sum_{i=0}^{255} \frac{1}{256} \cdot \log_2 256 = 256 \cdot \frac{1}{256} \cdot \log_2 256 = 8$$

- Fig. b shows the histogram of another image, in which 1/3 of the pixels are rather dark and 2/3 of them are rather bright. The entropy of this image is

$$\begin{aligned} \eta &= \frac{1}{3} \cdot \log_2 3 + \frac{2}{3} \cdot \log_2 \frac{3}{2} \\ &= 0.33 \times 1.59 + 0.67 \times 0.59 = 0.52 + 0.40 = 0.92 \end{aligned}$$

- The entropy is greater when the probability distribution is flat and smaller when it is more peaked

Histograms for two gray-level images. a Uniform distribution; b A sample binary image